

This electronic thesis or dissertation has been downloaded from the King's Research Portal at <https://kclpure.kcl.ac.uk/portal/>



## Sequelae of *Bacteroides fragilis* infection and carriage

Blandford, Lucy Emily

*Awarding institution:*  
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

### END USER LICENCE AGREEMENT



**Unless another licence is stated on the immediately following page** this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. <https://creativecommons.org/licenses/by-nc-nd/4.0/>

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

### Take down policy

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# **Sequelaes of *Bacteroides fragilis* infection and carriage**

Thesis submitted for the  
**Degree of Doctor of Philosophy (PhD)**

**LUCY EMILY BLANDFORD**

Department of Mucosal and Salivary Biology  
Dental Institute  
Kings College London

**2018**

## CONTENTS

TABLE OF FIGURES .....	8
ABSTRACT .....	12
ACKNOWLEDGEMENTS .....	14
LIST OF ABBREVIATIONS.....	15
Chapter 1 : INTRODUCTION.....	17
1.1 The Colon.....	17
1.1.1 Macroscopic Anatomy .....	17
1.1.2 Histology .....	18
1.1.3 The Gut microbiome .....	19
1.2 Inflammatory Bowel Disease (IBD) .....	20
1.2.1 What is inflammation? .....	20
1.2.2 Clinical aspects of Inflammatory Bowel Disease (IBD).....	21
1.2.3 Importance of the intestinal barrier in IBD .....	23
1.2.4 Interleukin-10 (IL-10) .....	26
1.2.5 Gut microbiome involvement .....	27
1.3 Cancer.....	29
1.3.1 What is cancer? .....	29
1.3.2 Cell Cycle and DNA damage .....	31
1.4 Colorectal Cancer (CRC) .....	32
1.4.1 Adenomatous polyps and colorectal cancer .....	32
1.4.2 Genetic pathways of CRC .....	34
1.4.3 Genomic instability.....	34
1.4.4 Importance of APC and $\beta$ -catenin signalling in CRC.....	36
1.4.5 Colonoscopy .....	37

## Chapter 1.1.1 Macroscopic Anatomy

1.5 Bacterial pathogenesis and virulence factors .....	37
1.5.1 Bacterial toxins.....	38
1.5.2 <i>B. fragilis</i> – an example of a pathobiont .....	39
1.6 Microbial-induced cancer .....	39
1.6.1 HPV - a viral example .....	40
1.6.2 Bacterial mechanisms of cell cycle modulation .....	40
1.6.3 Oncogenic modulation of host cell signalling by bacteria .....	42
1.6.4 <i>Helicobacter pylori</i> - a bacterial example.....	43
1.6.5 Modulation of $\beta$ -catenin signalling .....	45
1.7 CRC microbiome.....	47
1.7.1 Hypotheses of microbiota-induced CRC .....	47
1.7.2 <i>Escherichia coli</i> toxins and CRC.....	48
1.7.3 <i>Streptococcus gallolyticus</i> subsp. <i>gallolyticus</i> .....	49
1.7.4 <i>Fusobacterium nucleatum</i> .....	50
1.7.5 <i>Bacteroides fragilis</i> toxin (BFT).....	51
1.8 Aims .....	54
1.8.1 Why <i>B. fragilis</i> ? .....	54
1.8.2 Objectives .....	54
Chapter 2 : MATERIALS AND METHODS .....	56
2.1 Sample Collection .....	56
2.1.1 Subject Recruitment and Ethics Statement .....	56
2.1.2 Sample Collection and Processing .....	56
2.2 DNA Processing.....	68
2.2.1 DNA Extraction from Intestinal Biopsies.....	68
2.2.2 DNA Quantification .....	68
2.3 End-point Polymerase Chain Reaction (PCR).....	68

## Chapter 1.1.1 Macroscopic Anatomy

2.3.1 Primers.....	68
2.3.2 Polymerase Chain Reaction (PCR) .....	68
2.3.3 Colony PCR .....	69
2.3.4 DNA Agarose Gel Electrophoresis .....	69
2.3.5 PCR Amplicon Purification and Sequencing .....	69
2.4 Protein techniques .....	70
2.4.1 Protein Extraction from Intestinal Biopsy Samples.....	70
2.4.2 Total protein quantification.....	70
2.4.3 Enzyme-Linked Immunosorbant Assay (ELISA) .....	70
2.5 Microbiology .....	71
2.5.1 Microbiology.....	71
2.5.2 Culturing Conditions .....	72
2.5.3 Plate wash method to detect bacterial isolates .....	73
2.6 Polysaccharide A (PSA) Promoter Orientation .....	73
2.7 Statistical Analysis .....	74
Chapter 3 : END-POINT POLYMERASE CHAIN REACTION DEVELOPMENT .....	75
3.1 Introduction and Aims .....	75
3.1.1 Methods of bacterial identification and characterisation.....	75
3.1.2 Polymerase Chain Reaction .....	76
3.2 Method .....	78
3.2.1 DNA sequence retrieval and PCR primer design .....	78
3.2.2 Primer specificity testing .....	79
3.2.3 Annealing Temperature Optimisation .....	79
3.2.4 Limit of Detection .....	80
3.3 Results .....	81
3.3.1 DNA Extraction and Quality .....	81

## Chapter 1.1.1 Macroscopic Anatomy

3.3.2 Primer Design and Specificity Testing .....	83
3.3.3 Annealing Temperature Optimisation .....	95
3.3.4 PCR Inhibitors and Limit of Detection .....	96
3.3.5 Results Summary.....	97
3.4 Discussion .....	98
3.4.1 DNA template.....	98
3.4.2 Detection methods .....	99

## Chapter 4 : PREVALENCE OF BACTERIA AND GENES OF INTEREST IN PATIENT SAMPLES .....

101

4.1 Introduction .....	101
<b>4.2 Results</b> .....	101
4.2.1 Clinical results.....	101
4.2.2 Extracted DNA .....	102
4.2.3 Prevalence of <i>B. fragilis</i> , <i>bft</i> and <i>ubb</i> .....	103
4.2.4 Prevalence of <i>Fusobacterium</i> and <i>fadA</i> .....	107
4.2.5 Age and PCR positivity .....	107
4.2.6 Previous polyps/cancer.....	109
4.2.7 Individuals with multiple tumours .....	109
4.2.8 Cancer patients.....	110
4.2.9 Culture-based identification utilizing selective media .....	111
4.2.10 Bioinformatic data mining .....	113
4.3 Discussion .....	113
4.3.1 <i>B. fragilis</i> .....	113
4.3.2 <i>Bft</i> .....	113
4.3.3 <i>Ubb</i> .....	115
4.3.4 <i>Fusobacterium</i> , <i>fadA</i> and CRC.....	115

## Chapter 1.1.1 Macroscopic Anatomy

4.3.5 <i>Fusobacterium</i> and IBD .....	116
4.3.6 <i>B. fragilis</i> , <i>bft</i> and IBD .....	117
Chapter 5 : ANALYSIS OF BACTEROIDES FRAGILIS POLYSACCHARIDE A IN VIVO .....	118
5.1 Introduction .....	118
5.1.1 Antigenic and phase variation.....	118
5.1.2 Capsular Polysaccharide (PS) biosynthesis and transcription .....	119
5.1.3 Polysaccharide A (PSA) .....	120
5.1.4 Restriction-modification systems .....	121
5.2 Results .....	122
5.3 Polysaccharide A (PSA) promoter orientation .....	122
5.3.1 Restriction Digest optimization .....	122
5.3.2 Identification of a Single Nucleotide Polymorphism (SNP) in <i>UpaY</i> .....	123
5.3.3 PSA promoter orientation separated by biopsy location.....	126
5.3.4 PSA promoter orientation separated by patient groups .....	127
5.4 Interleukin 10 (IL-10) .....	129
5.4.1 Serum .....	129
5.4.2 Tissue lysate .....	130
5.4.3 Correlation analyses .....	133
5.4.4 Summary of Results .....	134
5.5 Discussion .....	135
5.5.1 PSA promoter orientation in vivo .....	135
5.5.2 Identification of a SNP .....	137
5.5.3 Serum IL-10 .....	137
5.5.4 Tissue IL-10 .....	138
5.5.5 PSA promoter orientation is not correlated with IL-10 concentration .....	139
5.5.6 The biological significance of capsular phase variation .....	139

Chapter 1.1.1 Macroscopic Anatomy	
5.5.7 Future research.....	140
Chapter 6 : GENERAL DISCUSSION.....	142
6.1.1 Polyposis.....	143
6.1.2 Biogeography.....	143
6.1.3 <i>Fusobacterium</i> and <i>F. nucleatum</i> taxonomy .....	144
REFERENCES.....	146



## TABLE OF FIGURES

Figure 1-1 Colon Anatomy.....	18
Figure 1-2 Intestinal mucosal barrier.....	24
Figure 1-3 IL-10 signalling pathway (Berti <i>et al.</i> , 2017) .....	27
Figure 1-4 Colonic tumourigenesis (adapted from Hazewinkel and Dekker 2011) .....	33
Figure 1-5 Types of Genomic Instability in Colorectal Cancer .....	35
Figure 1-6 Bacterial impairment of cell cycle regulation and DNA damage response .....	41
Figure 1-7 Bacterial modulation of host cell signalling .....	43
Figure 1-8 Pathogenic mechanisms of <i>B. fragilis</i> toxin (BFT) .....	53
Figure 3-1 Polymerase Chain Reaction .....	77
Figure 3-2 Assessment of DNA Quality from DNA Extracts .....	82
Figure 3-3 Amplification of bacterial 16S rRNA DNA from patient biopsy samples using universal bacterial primers (27F, 1292R) .....	83
Figure 3-4 Primer Specificity Testing: .....	84
Figure 3-5: Primer Design: <i>B. fragilis</i> .....	85
Figure 3-6 Evolutionary relationships of <i>Bacteroides</i> 16S rRNA gene and closely related <i>Bacteroidales</i> taxa .....	86
Figure 3-7 Primer Design and Specificity: <i>Fusobacterium</i> genus .....	88
Figure 3-8 Evolutionary relationships of <i>Fusobacterium</i> 16S rRNA gene and closely related taxa .....	89
Figure 3-9 Primer Design and Specificity Testing: <i>F. nucleatum</i> group.....	90
Figure 3-10 Primer Design and Specificity Testing: <i>F. nucleatum</i> subspecies <i>nucleatum</i> .....	91
Figure 3-11 Primer Design and Specificity Testing: <i>B. fragilis</i> toxin gene ( <i>bft</i> ) .....	93
Figure 3-12 Primer Design and Specificity Testing: <i>B. fragilis</i> ubiquitin gene ( <i>ubb</i> ).....	94
Figure 3-13 Primer Specificity Testing: <i>Fusobacterium fadA</i> .....	95
Figure 3-14 Annealing temperature optimisation: <i>B. fragilis</i> .....	96
Figure 3-15 Determination of PCR inhibition using spiked biopsy DNA: <i>ubb</i> .....	97
Figure 4-1 Age of individuals recruited to each patient group .....	102
Figure 4-2 Percentage of individuals (prevalence) PCR-positive for <i>B. fragilis</i> , <i>bft</i> and <i>ubb</i> in biopsy samples.....	104
Figure 4-3 Percentage of individuals positive for <i>Fusobacterium/FadA</i> .....	108

Figure 4-4 Comparison of age of individuals separated by <i>bft</i> (A) and <i>Fusobacterium</i> (B) positivity..	108
Figure 4-5 Percentage of individuals positive for gene targets: Comparison of alternative patient groups.....	110
Figure 4-6 Prevalence of <i>bft</i> and <i>ubb</i> in metagenomic studies .....	113
Figure 5-1 Schematic of the PS promoter and biosynthetic locus .....	120
Figure 5-2 Mechanism of immune cell activation .....	121
Figure 5-3 Optimisation of Restriction Digest Time for the PSA promoter orientation assay.....	123
Figure 5-4 Example Restriction Digest DNA Gel Electrophoresis .....	124
Figure 5-5 Polysaccharide A (PSA) Promoter Region DNA Sequences from Different <i>B. fragilis</i> strains .....	125
Figure 5-6 Schematic of Restriction digest fragments of the Polysaccharide A (PSA) promoter regions with extra <i>SspI</i> restriction site.....	126
Figure 5-7 Percentage of <i>B. fragilis</i> cells with the Polysaccharide A (PSA) promoter in the 'On' orientation: comparison between colonic biopsy samples .....	126
Figure 5-8 Percentage of <i>B. fragilis</i> cells with the Polysaccharide A (PSA) promoter in the 'On' orientation: comparison amongst patient groups .....	127
Figure 5-9 Individuals with <i>bft</i> positive strains have a significantly lower percentage of <i>B. fragilis</i> with the PSA promoter in the 'ON' position .....	128
Figure 5-10 Frequency distributions of serum and tissue IL-10 concentration do not follow a normal distribution .....	129
Figure 5-11 Serum IL-10 concentration separated by patient groups or between <i>B. fragilis</i> -positivity .....	130
Figure 5-12 Tissue IL-10 concentration amongst biopsy sites .....	131
Figure 5-13 Tissue IL-10 concentration amongst patient groups.....	131
Figure 5-14 Tissue IL-10 concentration as separated by the presence of BF or BFT.....	132
Figure 5-15 Correlation analysis of serum IL-10 or tissue IL-10 and <i>B. fragilis</i> PSA promoter orientation .....	133
Figure 5-16 Scatter graphs to identify possible correlation of variables with age of the individual ...	134

## TABLE OF TABLES

Table 1-1 Differentiating factors between Crohn's Disease and Ulcerative Colitis .....	23
Table 1-2 Definitions of types of cellular adaptation (Weinberg, 2014) .....	29
Table 1-3 Hanahan and Weinberg Hallmarks of Cancer .....	30
Table 1-4 Summary of adenomatous polyp subtypes (Bosman <i>et al.</i> , 2010) .....	33
Table 1-5 Bacterial pathogens and bacterial dysbiosis epidemiologically linked to a specific cancer type .....	40
Table 1-6 <i>E. coli</i> findings in human colorectal cancer (CRC) samples .....	49
Table 2-1 Healthy patient cohort .....	58
Table 2-2 Inflammatory Bowel Disease Patient Cohort .....	61
Table 2-3 Previous polyps/cancer patient cohort .....	62
Table 2-4 Newly diagnosed polyp/cancer patient cohort .....	64
Table 2-5 Patients assigned multiple cohorts .....	67
Table 2-6 Table of Bacterial Reference Strains .....	71
Table 2-7 Table of <i>Bacteroides fragilis</i> Reference Strains .....	72
Table 2-8 Table of <i>Fusobacterium nucleatum</i> Reference Strains .....	72
Table 2-9 PCR digestion fragment sizes .....	74
Table 2-10 Polysaccharide A Promoter Region Primers .....	74
Table 3-1 Primer Characteristics recommended for an efficient PCR .....	79
Table 3-2 Primers designed and used for specificity testing .....	80
Table 3-3 Primers and optimised PCR Conditions used for .....	97
Table 4-1 Summary of individuals and samples recruited for the study .....	101
Table 4-2 Numbers of individuals positive for <i>B. fragilis</i> , <i>bft</i> and <i>ubb</i> .....	104
Table 4-3 Clinical characteristics of <i>ubb</i> positive individuals .....	105
Table 4-4 Clinical characteristics of <i>bft</i> positive individuals .....	106
Table 4-5 Numbers of individuals positive for <i>Fusobacterium</i> and <i>fadA</i> .....	107
Table 4-6 PCR results of individuals with histologically defined cancer .....	111
Table 4-7 Summary of plate wash PCR results from caecal effluent samples compared with biopsy PCR .....	112

<b>Table 4-8 Summary of the literature on <i>bft</i> epidemiology.....</b>	<b>115</b>
<b>5-1 Common mechanisms of antigenic variation.....</b>	<b>118</b>
<b>Table 5-2 Summary of p values for significant differences between the variables tested (PSA promoter orientation, serum IL-10 concentration and tissue IL-10 concentration) and presence of bacterial taxa or virulence genes as detected by PCR .....</b>	<b>135</b>
<b>Table 5-3 Summary of p values for pairwise significant differences between patient groups for all of the variables tested (PSA promoter orientation, serum IL-10 concentration and tissue IL-10 concentration) .....</b>	<b>135</b>
<b>Table 5-4 Summary of literature measuring serum IL-10 Levels in IBD and CRC .....</b>	<b>138</b>

## ABSTRACT

*B. fragilis* is considered an opportunistic pathogen, often isolated from abdominal abscesses, bloodstream infections and peritonitis. *B. fragilis* can produce multiple capsular polysaccharides including an immunomodulatory, zwitterionic, polysaccharide A (PSA) capable of stimulating anti-inflammatory interleukin-10 (IL-10) production. Conversely, some strains can produce a putative carcinogenic toxin, the metalloprotease *Bacteroides fragilis* toxin (BFT). BFT has been demonstrated to promote colonic cell proliferation and DNA damage in mammalian cell culture and animal models. An additional putative *B. fragilis* virulence factor present in select strains is a eukaryotic-like ubiquitin protein (BfUbb). BfUbb is capable of interfering with the host ubiquitination cascade this protein but the consequence of this on host health is unknown.

This study describes the robust design and validation of PCR-based assay to target specific bacterial taxa and putative virulence genes. The PCR assay was subsequently used to determine prevalence in a collection of gastrointestinal tissue samples from individuals with and without disease. The *bft* gene was found to have a significantly higher prevalence in individuals newly diagnosed with polyps/cancer compared with a healthy patient group. This finding further points towards the importance of BFT in colonic tumorigenesis. Contrary to previous CRC literature the prevalence of *Fusobacterium* and *fadA* were not significant in the cohorts investigated in this study. Colonic location and histological type of *Fusobacterium*-positive tumours did not result in any significant associations, but the trends observed support previous suggestions of an association between *Fusobacterium* species and right-sided colon cancer.

Presented here is the first reported determination of *B. fragilis* capsular PSA promoter orientation *in vivo*. Furthermore, individuals with IBD had a significantly lower percentage of the *B. fragilis* population PSA orientated on in comparison with a healthy cohort. Similarly, *bft*-positivity was significantly associated with a lower proportion of the PSA promoter orientated on.

In conclusion, overall the results presented indicate that the common gastrointestinal species, *B. fragilis*, can have wide ranging effects on gastrointestinal health. Between-strain differences and within-strain antigenic variation were shown to have significant associations with patient

### Chapter 1.1.1 Macroscopic Anatomy

populations and argue for a gene-centric, and not taxonomic-centric, approach to microbiome research.

## **ACKNOWLEDGEMENTS**

Firstly, I would like to thank my first supervisor Professor Alistair Lax for his expertise and guidance throughout my PhD. I have developed a more critical thinking style and become more thorough in my approach to research under his supervision. Maybe one day I will argue back!

Secondly, I would to express my appreciation to Professor Jeremy Sanderson for his assistance in establishing the patient sample collection and valuable clinical input. A special thank you to all of the staff at Guy's and St. Thomas' Endoscopy department who were extremely helpful and friendly during my time recruiting patients and collecting samples. Dr Ikram Nasr and Dr Emma Johnson, in particular gave their time and knowledge to help me with the study.

I would like to acknowledge my third supervisor Professor William Wade who helped with PCR methodology, microbiology input and the quickest response to emails ever!

I would like to thank all at ImmBio for aiding with my PhD funding and teaching me many valuable lab skills during my time working for them. I am very grateful to Camilo for his infectious enthusiasm for science and pushing me to do a PhD – here is the outcome!

A massive thank you to everyone within the salivary and microbiology group and labs who have been very helpful in answering my questions, providing guidance for experiments but most of all for their friendship. A special mention goes to my triplets Matt Blakeley and Arshiya Banu for gossiping sessions, making me laugh and memorable pub trips! Thank you to Steve Gilbert for keeping the microbiology lab tidy, organised and always fully stocked with all the lab consumables I could wish for!

Lastly, thank you to my parents and sister for their support during my PhD. Having to explain my research in layman's terms to them has enhanced my view of the overall picture and reasons behind the research. They are still not sure exactly what I've been researching so here it is - please read every word!

## LIST OF ABBREVIATIONS

### A

---

AC Ascending Colon

ADAM A Disintegrin and Metalloprotease

APC Adenomatous Polyposis Coli

AT Away from tumour (~10 cm away)

### B

---

BFT *Bacteroides fragilis* toxin

BHI Brain-Heart Infusion broth

BLAST Basic Local Alignment Search Tool

### C

---

CagA Cytotoxin-associated gene-A

CARD9 Caspase Recruitment Domain  
family member 9

CD Crohn's Disease

CDK Cyclin-dependent Kinase

CDT Cytolethal Distending Toxin

CE Caecal Effluent

CIMP CpG island Methylator Phenotype

Cif Cycle-inhibiting factor

CIN Chromosomal Instability

CK1 Casein Kinase 1

CNF Cytolethal Necrotizing Factor

COPD Chronic Obstructive Pulmonary  
Disease

CRC Colorectal Cancer

### D

---

DC Descending Colon

DDR DNA Damage Response

DSS Dextran Sodium Sulfate

### E

---

EGFR Epidermal Growth

Factor Receptor

ELISA Enzyme-Linked Immunosorbant  
Assay

ETBF Enterotoxigenic *Bacteroides fragilis*

### F

---

FAP Familial Adenomatous polyposis

FISH Fluorescent *in situ* hybridization

### G

---

GI tract Gastrointestinal tract

GSK-3 Glycogen synthase kinase 3

GWAS Genome wide association study

### H

---

HIOEC Human Immortalized Oral Epithelial  
Cell

HMP Human Microbiome Project

HPV Human Papilloma Virus

### J

---

JCM Japan Collection of Microorganisms

### I

---

IBD Inflammatory Bowel Disease

IgA Immunoglobulin A

IL-10 Interleukin 10

IL-17 Interleukin 17

### K

---

KEGG Kyoto Encyclopedia of Genes and  
Genomes

### L

---

LLO Listeriolysin O



LPS Lipopolysaccharide

## **M**

---

MC Microcapsule

MCR Mutation Cluster Region

MEGA Molecular Evolutionary Genetic Analysis

MHC Major Histocompatibility Complex

MMP Matrix Metalloproteases

MMR Mismatch Repair

MSI Microsatellite Instability

MUSCLE Multiple Sequence Comparison by Log-Expectation

## **N**

---

NOD2 Nucleotide-binding Oligomerization Domain-containing protein 2

NT Next to tumour (~2 cm away)

## **O**

---

OGD Oesophago-gastroduodenoscopy

OMV Outer Membrane Vesicle

OT On tumour

## **P**

---

PAI Pathogenicity island

PCR Polymerase Chain Reaction

PI-3K Phosphoinositide 3-kinase

PIP3 Phosphatidylinositol-3,4,5-Triphosphate

PS Polysaccharide

PSA Polysaccharide A

PSC Primary Sclerosing Cholangitis

## **R**

---

RB Retinoblastoma

RFLP Restriction Fragment Length

Polymorphism

ROS Reactive Oxygen Species

RVFV Rift Valley Fever Virus

## **S**

---

SASP Senescence-associated Secretory Phenotype

SCFA Short Chain Fatty Acids

SH2 SRC homology 2 domain

SMO Spermine Oxidase

SMT Somatic Mutation Theory

SNP Single Nucleotide Polymorphism

SPF Specific pathogen free

STAT3 Signal Transducers and Activators of Transcription 3

## **T**

---

T4SS Type IV secretion system

T6SS Type VI secretion systems

TAMP Tight-junction Associated Marvel domain-containing Protein

TCR T-cell Receptor

TLR2 Toll-like Receptor 2

T<sub>m</sub> Melting temperature

TNF  $\alpha$  -Tumour Necrosis Factor- $\alpha$

TOFT Tissue Organisation Field Theory

## **U**

---

UC Ulcerative Colitis

## **V**

---

VacA Vacuolating cytotoxin

VlsE Vmp-like sequence

## **Z**

---

ZO Zonula Occludins

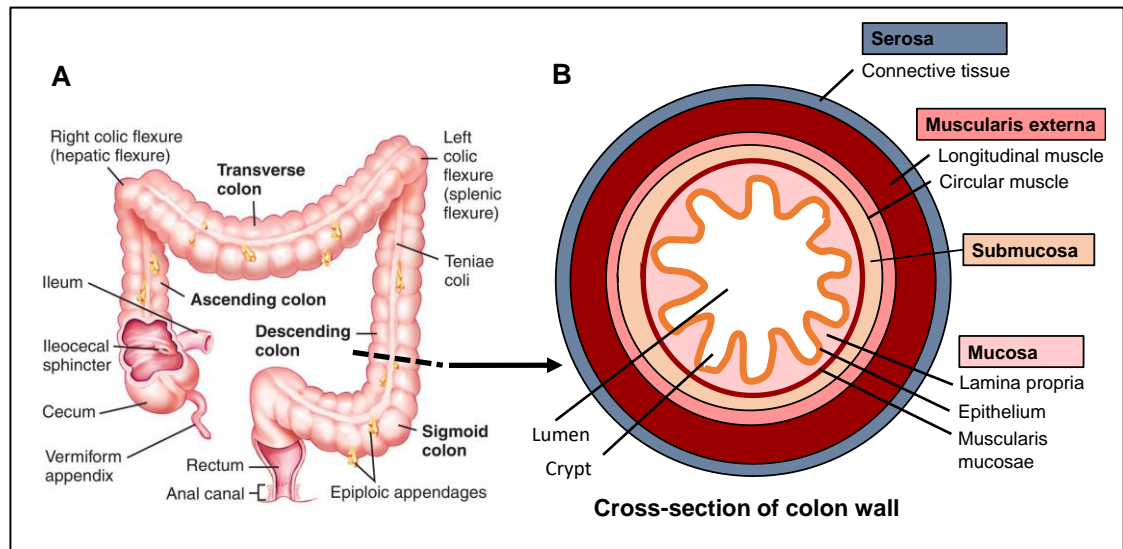
## **Chapter 1 : INTRODUCTION**

### **1.1 The Colon**

The main functions of the colon are considered to be water and electrolyte absorption and microbiome metabolism. The majority of nutrient absorption occurs within the small intestine however this is often accompanied by the release of water into the lumen. An important role of the colon is therefore to conserve this water and to concentrate waste faecal matter. Approximately 90% of water that is released through the ileocaecal valve into the colon is recovered (Debonigie and Phillips, 1978).

#### **1.1.1 *Macroscopic Anatomy***

The distal region of the alimentary canal is the large intestine, which ranges from the ileocaecal valve to the anus; including the colon and rectum. The colon originates at the caecum and is divided from the small intestine by the ileocaecal valve. The caecum contains a worm-shaped outpouching known as the vermiform appendix, notorious for being a potential life-threatening site of inflammation. Proceeding distally along the colon is the ascending colon (on the right side of the abdomen) which bends at the hepatic flexure to become the transverse colon. The transverse colon bends at the splenic flexure into the descending colon (left side of the abdomen). Distal to the descending colon is the S-shaped curvature of the sigmoid colon before the colon terminates at the rectum and anus (Figure 1-1A).



**Figure 1-1 Colon Anatomy**

(A) Macroscopic anatomy of the colon adapted from (Ellis, 2010).

(B) A cross-sectional representation of the layered colon wall

### 1.1.2 Histology

From the outside of the colonic wall to the inside luminal space the colonic wall has four distinct layers (Figure 1-1B):

- I. Serosa
  - II. Muscularis externa
  - III. Submucosa
  - IV. Mucosa
- I. The serosa is the outer lining of the colon and rectum and is composed of connective tissue. This layer is lined by mesothelial cells and their associated basement which secrete fluid to allow for movement of the organ within the body cavity. Pouches of fat or appendices epiploicae are located intermittently attached to the serosa.
  - II. The muscularis externa is made up of two layers of muscle cells, an outer longitudinal muscle layer and an inner (closer to the lumen) sheet of circular muscle. Parasympathetic nervous tissue (myenteric plexus) is situated between the two layers innervating the muscle for peristalsis.

- III. The submucosa is an area of loose connective tissue also containing neurones, blood vessels and, scavenging immune cells.
- IV. The mucosa comprises the epithelial cell layer, lamina propria (connective tissue) and the muscularis mucosae (a thin layer of smooth muscle cells). The mucosal epithelium surface of the colon is smooth without villi, only regularly arranged colonic glands. These glandular structures are known as the crypts of Lieberkühn and extend down towards the muscularis mucosae (Figure 1-1B). Cell types that make up this epithelial layer are predominantly columnar absorptive enterocytes but there are also many goblet cells. The goblet cells secrete mucus which serves to protect the colonic epithelium and aid the passage of faecal matter. Proliferating stem cells are located at the base of the crypt which migrate upwards as they differentiate. Supporting the epithelium and surrounding the crypts is the lamina propria. This layer is predominantly loose connective tissue containing many fibroblasts but also holds a large population of immune cells (lymphocytes) and neurovascular elements (Kerr, 2010).

### **1.1.3 The Gut microbiome**

In more recent years the significance of the influence of the colonic microbiome on human health has been appreciated and consequently is now considered a valuable function of the colon and the body as a whole. The metagenomic sequencing projects, Human Microbiome Project (HMP) (Group *et al.*, 2009) and MetaHIT (Li *et al.*, 2014) revealed a vast genomic catalogue within this 'second genome' and a surprising amount of diversity in species composition between individuals, although gene function is relatively conserved (Group *et al.*, 2009).

A significant function of the microbiome is bacterial fermentation of dietary fibre into short chain fatty acids (SCFAs), typically acetate, propionate and butyrate. Humans lack the enzymes to break down complex carbohydrates which instead are metabolised by anaerobic bacteria in the caecum and colon. SCFAs can then be used as an energy source by the host, particularly in the

### Chapter 1.2.1 What is inflammation?

maintenance of colonocytes. Microbes within the gut are also able to synthesise vitamins and essential amino acids of benefit for host nutrition (den Besten *et al.*, 2013).

The density of microbes increases distally along the GI tract. In the small intestine the concentration of microbes is limited through host mechanisms such as the secretion of immunoglobulin A (IgA) (Macpherson and McCoy, 2015) and antimicrobial peptides (Vaishnava *et al.*, 2011). This limited microbial component enables nutrient absorption while in the colon density is increased as oxygen concentration decreases and SCFA production is encouraged. The colonic epithelium is covered in a layer of mucus which can be subdivided into an inner dense layer and a luminal-exposed less dense layer. It was originally thought based on fluorescent *in situ* hybridization (FISH) techniques that this mucus layer was effectively sterile (van der Waaij *et al.*, 2005). In recent years there has been mounting evidence for the presence of a crypt-associated microbiota. Techniques such electron microscopy, laser microdissection and methodological improvements to conserve the mucus layer during histological fixation and washing have revealed a community of primarily aerobic *Acinetobacter* sp. but also the common GI commensal *Bacteroides fragilis* (Donaldson *et al.*, 2016, Lee *et al.*, 2013).

## 1.2 Inflammatory Bowel Disease (IBD)

IBD is of unknown aetiology but is likely to be multifactorial through a combination of genetics, the host immune system and environmental factors. An over reactive immune reaction to the commensal gut microbiota and disruption to the intestinal mucosal barrier are thought to be key in IBD pathogenesis (Scaldaferri and Fiocchi, 2007).

### 1.2.1 What is inflammation?

The process of inflammation describes the immunovascular response to tissue injury. A change in normal tissue homeostasis may be a consequence of harmful biological, chemical or physical stimuli. The host protects the tissue against these stimuli and resolves tissue damage via the induction of inflammatory mechanisms (Ahmed, 2011). At a macroscopic level inflammation is classically characterised by redness, heat, pain and swelling. Histologically inflammation is accompanied by an increased blood supply, increased permeability of blood vessels and infiltration of leukocytes into the tissue. Altogether this allows for the recruitment of relevant immune cells to clear any infection and initiate wound repair mechanisms. The stages of the

## Chapter 1.2.2 Clinical aspects of Inflammatory Bowel Disease (IBD)

acute inflammatory response are initiation of inflammation to rapidly recruit leukocytes to the injured tissue, amplification of the response depending on the extent of the injured tissue, destruction and elimination of the harmful stimuli and finally termination of the response including healing of damaged tissue (Ward and Lentsch, 1999). In some cases this acute inflammation is not resolved resulting in chronic inflammation which may include persistence of a pathogen, tissue damage, fibrosis, autoimmune conditions and even cancer (Kumar *et al.*, 1997).

### **1.2.2 Clinical aspects of Inflammatory Bowel Disease (IBD)**

IBD is characterised by chronic inflammation in the gastrointestinal tract and is comprised of two main types, Crohn's disease (CD) and Ulcerative colitis (UC) (Fakhoury *et al.*, 2014). The two subtypes differ primarily on the GI tract location and extent of inflammation. Main differentiating features between UC and CD are that in CD inflammation is transmural (across the entire wall of the bowel) and can be present anywhere along the length of the G.I tract. In contrast, UC inflammation is limited to the mucosal layer and often starts in the rectum spreading proximally, also known as pancolitis when the inflammation extends along the length of the colon. Diagnosis is made by a combination of clinical, laboratory, endoscopic and radiographic observations. Due to the similarities between CD and UC, diagnosis of the disease subtype is not always straightforward and sometimes is classed as indeterminate colitis. Approximately 10-15% of individuals diagnosed with IBD (Guindi and Riddell, 2004) are categorised as having indeterminate colitis. Determination of the disease subtype influences patient management and therefore correct diagnosis may determine the responsiveness to treatment. Differentiating factors between CD and UC are summarised in Table 1-1 and discussed below.

#### **1.2.2.1 Histopathology**

Characteristic histological features of UC disease are mucin depletion, neutrophil infiltrate in the lamina propria and altered crypt architecture (Geboes, 2003). Crypt histopathology includes transepithelial migration of neutrophils, and sometimes eosinophils, into the mucosal epithelium known as cryptitis and is considered a measure of active disease. Crypt abscesses may also form in which this infiltrate migration extends to the lumen. Fibrotic regions of the mucosa and submucosa become more common with repeated tissue injury and healing which may result in a narrowed and shortened colon. Over time UC affected colonic crypts may become shortened,

### Chapter 1.2.2 Clinical aspects of Inflammatory Bowel Disease (IBD)

branched or bifid (with double lumina). Other long-term artefacts of UC are repeated healing of mucosal ulcers may lead to elevated sessile nodules known as pseudopolyps (Kelly and Gabos, 1987).

In early CD ulcers are often small, multiple superficial lesions, known as aphthoid ulcers which tend to be found in areas of the mucosa overlying lymphoid follicles (Magro *et al.*, 2013). Multiple ulcerated regions surrounded by unaffected areas are known as 'skip lesions' and over time may converge to form deep, discrete ulcers. This combination of longitudinal and transverse ulceration gives rise to the classic histological 'cobblestone appearance'. Extensive transmural ulcers may pass through the muscularis and create a connecting pathway, or fistula, to another organ or outside of the body.

Microscopic appearance of CD is characterised by epithelioid granulomas which are activated macrophages resembling epithelial cells. Granulomas may be associated with blood vessels, or may be limited to inflammatory cell infiltrate. Submucosal features of CD are lymphoid aggregates, nerve hyperplasia and over time large amounts of collagen. Collagenous material also present in the muscularis mucosae alongside increased smooth muscle cells, collagen, laminin, tenascin and abundant mast cells can result in narrowing, or stricturing, of the bowel wall. Strictures are a common complication of longstanding CD.

An additional type of intestinal inflammation is microscopic colitis which is characterised by the presence of watery diarrhoea, a normal appearance of the colon at endoscopy but with a distinct histological pattern.

**Table 1-1 Differentiating factors between Crohn's Disease and Ulcerative Colitis**

	Crohn's Disease	Ulcerative Colitis
<b>Location</b>	Anywhere - terminal ileum is the most usual location of inflammation	Colon - usually starts in the rectum and extends proximally
<b>Symptoms</b>	Peri-anal problems more common. Strictures and fistulae	Rectal bleeding more common
<b>Inflammation</b>	Can be segmental (skip lesions) and cobblestone appearance	Continuous area
<b>Histological features</b>	Epithelioid granulomas, fat wrapping, increased thickness of bowel wall	Mucin depletion, distorted crypt architecture; lamina propria cell infiltrate
<b>Pseudopolyps</b>	Less common but tend to be depressed	More common
<b>Tobacco</b>	Aggravating	Preventative

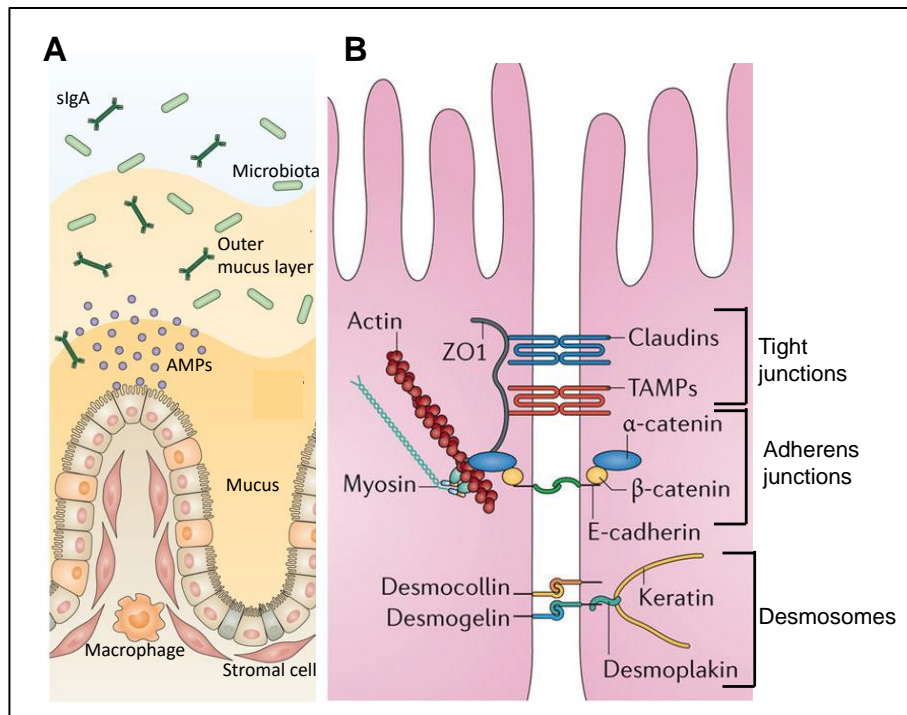
### **1.2.3 Importance of the intestinal barrier in IBD**

The intestinal mucosal barrier is a physical 'wall' comprised of specialised components whose role is to provide a shield from the potentially harmful luminal contents, predominantly microbial pathogens and environmental toxins. Barrier integrity is maintained by a number of components including specialised epithelial cell types, cell-cell junctions and the mucus constituents (Figure 1-2A). A variety of techniques including the lactulose/mannitol test (Fries *et al.*, 2005), confocal laser endomicroscopy (Buda *et al.*, 2014) and <sup>51</sup>Cr-ETDA absorption (Jenkins *et al.*, 1988) have shown an increased intestinal permeability in IBD. Increased paracellular permeability may allow for translocation of luminal contents to the basement membrane and induction of an inflammatory response (Chang *et al.*, 2017). The importance of these barrier-maintaining components, the consequence of intestinal barrier dysfunction, and the link with IBD is discussed below.



### Mucosal layer

The colonic epithelium is coated in mucus including a thick, viscous inner layer of the mucus and an outer thinner mucus layer (Figure 1-2A). Features of the intestinal mucus are its lubricant properties to reduce friction with passing luminal contents and its 'sticky' nature to trap ingested particles requiring removal (Cone, 2009, Kamphuis *et al.*, 2017). The bulk of mucus contents is formed from heavy glycosylated mucin oligomers with surface acting hydrophobic phospholipids (Lichtenberger, 1995).



**Figure 1-2 Intestinal mucosal barrier**

(A) Colonic intestinal barrier functions to separate luminal contents from the mucosal immune system. Taken from (Peterson and Artis, 2014)

(B) Adjacent epithelial cells with adjoining cell-cell junctions. Schematic depicts tight junctions, adherens junctions and desmosomes. Figure adapted from (Odenwald and Turner, 2017). TAMPs = Tight-junction associated marvel domain-containing proteins

Evidence of the importance of mucus layer in maintenance of the intestinal barrier can be observed in *Muc2*<sup>-/-</sup> mice that lack the gene required for synthesis of the most common intestinal mucin, mucin 2 (Van der Sluis *et al.*, 2006). Phenotypically these mice have altered crypt architecture and over time spontaneously develop intestinal colitis. Using the aforementioned *Muc2*<sup>-/-</sup> mouse model combined with treatment of a chemical inducer of colitis (dextran sulfate sodium (DSS)) severe colitis was induced (Petersson *et al.*, 2011). This more

### Chapter 1.2.3 Importance of the intestinal barrier in IBD

pronounced colitis, compared with DSS treatment only, resulted in bacterial translocation. Altogether the research suggests an outcome of a reduced mucus layer is that the barrier function is overridden with subsequent inflammation.

#### **Cell-cell junctions**

Another important component maintaining the intestinal barrier are cell-cell junctions between adjacent epithelial cells (Figure 1-2B). Multiple types of junction exist between cells that allow for intercellular communication and a selectively permeable barrier. Cell-cell junctions also help to maintain the intestinal barrier when epithelial cells are shed from the surface of the epithelium (Watson and Hughes, 2012).

Tight junctions are comprised of occludins, claudins, tricellulin and junctional adhesion molecule protein and have a role in selective ion permeability. The tight junction proteins interact with the cytoskeleton scaffold zonula occludens which itself is linked to F-actin (Figure 1-2B). Claudins are considered the backbone proteins of tight junctions and studies suggest this protein is key in tight junction regulated intestinal permeability. Studies have shown that claudin-7<sup>-/-</sup> mice develop colonic inflammation and increased paracellular permeability of small organic solutes (Tanaka *et al.*, 2015). Occludin protein expression is decreased at the site of inflamed mucosa in CD and UC while claudin-1 and zonula occludens (ZO) only have decreased expression at the edge of the inflamed site (Kucharzik *et al.*, 2001).

E-cadherin is a transmembrane protein that is important in cell-cell adhesion via the formation of Ca<sup>2+</sup>-dependent intra- and intercellular homophilic interactions between its' extracellular domains (Hatta *et al.*, 1988). These cadherin connections between neighbouring cells are the backbone of adherens junctions. The intracellular domain of E-cadherin interacts with  $\beta$ -catenin which is associated with  $\alpha$ -catenin and the actin cytoskeleton (Yamada *et al.*, 2005). In a DSS-induced mouse model of colitis intestinal deficiency of E-cadherin aggravates disease (Grill *et al.*, 2015). Genetic variants of the E-cadherin gene (*CDH1*) result in mis-localisation of the protein and are associated with CD (Muise *et al.*, 2009, van Heel *et al.*, 2003). In actively inflamed UC gene transcription but no protein expression of E-cadherin can be detected (Gassler *et al.*, 2001).

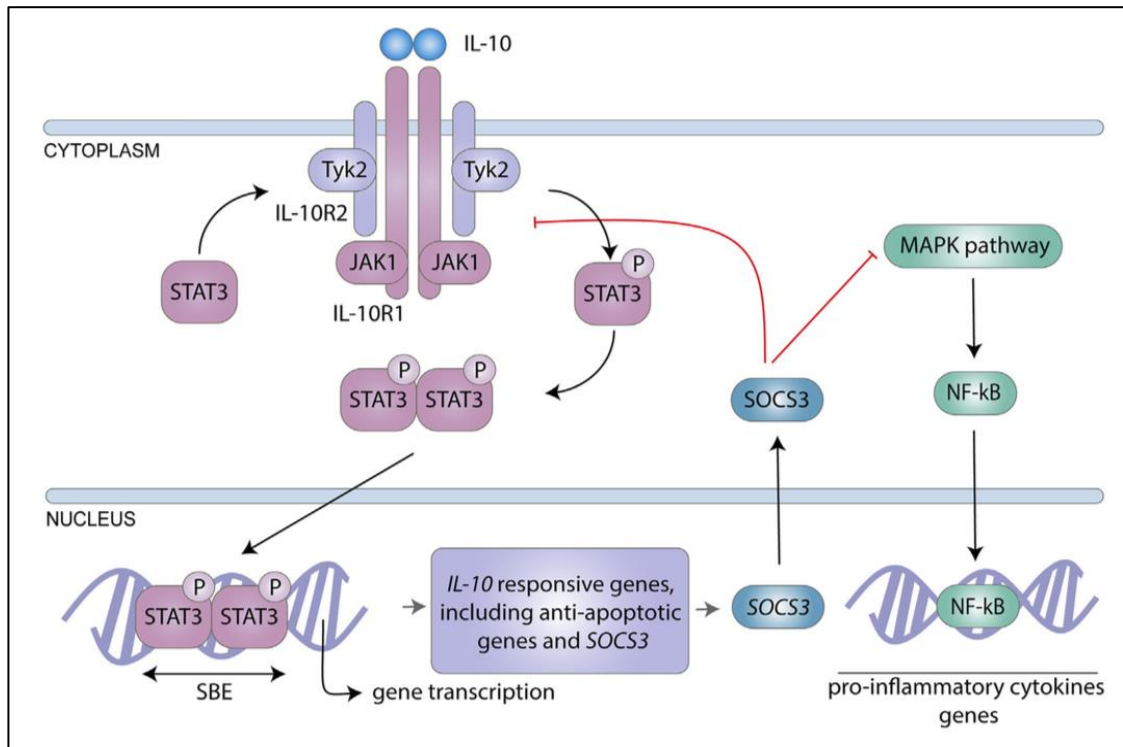
#### Chapter 1.2.4 Interleukin-10 (IL-10)

Desmosomes are strongly adhering junctional complexes linked to intermediate filaments and capable of resisting mechanical stress (Garrod and Chidgey, 2008). Little is known about their importance in intestinal homeostasis but loss of E-cadherin in intestinal epithelial cells results in loss of desmosomes (Schneider *et al.*, 2010).

#### **1.2.4 Interleukin-10 (IL-10)**

IL-10 is considered a potent anti-inflammatory cytokine capable of resolving inflammation. While leukocytes are the predominant cellular source of IL-10 it can also be produced by intestinal epithelial cells (Hyun *et al.*, 2015). Homodimeric IL-10 binds to heterotetrameric IL-10 receptor complex to activate IL-10 signalling. Ligand binding induces the associated Jak1 and Tyk2 tyrosine kinases to phosphorylate an intracellular portion of the IL-10 receptor. Phosphorylation of the receptor results in recruitment of signal transducer and activator of transcription 3 (STAT3) protein (Weber-Nordt *et al.*, 1996). Jak1 and Tyk2 also phosphorylate the newly associated STAT3, which leads to its homodimerisation, translocation to the nucleus and altered expression of IL-10 regulated genes (Donnelly *et al.*, 1999, Shouval *et al.*, 2014). Anti-inflammatory actions of IL-10 include inhibition of NF- $\kappa$ B signalling, with a subsequent decreased pro-inflammatory gene expression. Figure 1-3 summarises the key aspects of the IL-10 signalling pathway.

IL-10 is important for maintaining intestinal mucosal homeostasis and resolving inflammatory responses to commensal microorganisms. Evidence for this can be seen in IL-10<sup>-/-</sup> mice, which develop enterocolitis along the entire intestinal tract. Similarly, IL-10 receptor 1 knockout mice also develop spontaneous colitis (Spencer *et al.*, 1998). The importance of the microbiota in colitis is highlighted by studies undertaken in IL-10<sup>-/-</sup> mice which showed that under specific pathogen free (SPF) conditions IL10<sup>-/-</sup> develop local, attenuated disease (Kuhn *et al.*, 1993). Commensal *Bacteroides* species (Bloom *et al.*, 2011), *Helicobacter hepaticus* (Kullberg *et al.*, 1998) and other bacterial species have all been shown to initiate colitis in IL-10/receptor deficient mice but not in wild-type mice.



**Figure 1-3 IL-10 signalling pathway (Berti *et al.*, 2017)**

Binding of the IL-10 homodimer to the IL-10 receptor initiates phosphorylation of Tyk2 and JAK1. Stat3 is phosphorylated by the Tyk2 and JAK1 kinases resulting in its homodimerization. The activated Stat3 dimer translocates to the nucleus where it can act as a transcription factor inducing IL-10 responsive gene transcription. The SOCS3 protein negatively feeds back onto the signalling pathway and inhibits MAPK activated pro-inflammatory gene transcription.

In addition to the findings from mouse models a connection between IL-10 signalling and human IBD has been identified through genome wide linkage and association studies (GWAS) (Pigneur *et al.*, 2013), for example early-onset IBD with a severe phenotype which has been linked to mutations in the IL-10 receptor resulting in decreased STAT3 activation following IL-10 stimulation (Glocker *et al.*, 2009).

Currently the literature suggests interleukin-10 signalling is an important component in maintaining the intestinal barrier and aiding tolerance to the commensal microbiota and therefore is likely to be involved in the aetiology of IBD.

### 1.2.5 Gut microbiome involvement

A dysbiotic microbiome with decreased alpha-diversity, or species richness, has been observed in many studies of the IBD microbiota (Kostic *et al.*, 2014). Characteristics of the observed

#### Chapter 1.2.5 Gut microbiome involvement

changes are an increase in *Proteobacteria* (Gevers *et al.*, 2014) and a decrease in anaerobes and anti-inflammatory commensals e.g. *Faecalibacterium prausnitzii* (Sokol *et al.*, 2008). These studies are informative for knowledge of the IBD microbiome but do not distinguish between IBD-associated and/or causative microbes. Microbial metabolites and impaired microbiome sensing by the host immune system are mechanisms showing the significance of the microbiome in promoting intestinal inflammation.

The first gene to be linked to Crohn's Disease was Nucleotide-binding Oligomerization Domain-containing protein 2 (NOD2) (Hugot *et al.*, 2001). The NOD2 ligand is a muramyl dipeptide, a motif of bacterial peptidoglycan, and so this innate immune receptor is considered an intracellular bacterial sensor. GWAS have also linked the autophagy genes ATG16L1 (Hampe *et al.*, 2007) and IRGM (Parkes *et al.*, 2007) genes to CD susceptibility. Autophagy is a cellular recycling process but is also involved in clearance of pathogens so could indicate the involvement of an intracellular pathogen (Pareja and Colombo, 2013). NOD2 has subsequently been shown to recruit the ATG16L1 protein to the plasma membrane with mutant NOD2 failing to form effective bacterial clearing autophagosomes (Travassos *et al.*, 2010). CD-associated NOD2 and ATG16L1 are also associated with poor autophagy induction, bacterial trafficking, and antigen presentation in dendritic cells (Cooney *et al.*, 2010).

Caspase recruitment domain family member 9 (CARD9) is another IBD associated protein with links to the microbiota. IBD risk alleles exist in the caspase recruitment domain family member 9 (CARD9) gene (Beaudoin *et al.*, 2013) and CARD9<sup>-/-</sup> mice are more susceptible to colitis (Lamas *et al.*, 2016). CARD9 is linked with the intracellular sensing of pathogens via its association with the aforementioned NOD2 (Hsu *et al.*, 2007), induction of IL-22 (Bergmann *et al.*, 2017), fungal innate immunity (Gross *et al.*, 2006) and microbial anti-inflammatory aryl hydrocarbon receptor signalling (Lamas *et al.*, 2016).

Some microbial metabolites are able to positively influence the integrity of the intestinal barrier. The SCFA butyrate can enhance tight junction formation (Peng *et al.*, 2009) and SCFAs can signal through the regulatory T cell GPCR43 receptor to protect against intestinal inflammation (Smith *et al.*, 2013). More recently *Peptostreptococcus russellii* has been shown to increase

Muc2 producing goblet cells via the production of the small metabolite indoleacrylic acid (Wlodarska *et al.*, 2017) thereby increasing mucus production.

## 1.3 Cancer

### 1.3.1 What is cancer?

Cancer is a broad term used to describe many diseases originating from different cell types and anatomical sites. For this reason, to define cancer is a somewhat philosophical question but a key characteristic is uncontrolled cellular growth by cells that have acquired a cancerous, or malignant, phenotype. Cancer is a multistage process of monoclonal origin whereby a single cell acquires a malignant phenotype in a process known as cell transformation. Transformed cells proliferate indefinitely to form a mass of cells known as a tumour. Tumours are described as cancerous once they have evolved to invade surrounding, healthy tissue and sometimes have the ability to metastasise to other tissues. Indeed, in Colorectal Cancer (CRC) tumours are defined as cancerous if they invade into the bowel wall (Ruddon, 2007, Knowles and Selby, 2005).

**Table 1-2 Definitions of types of cellular adaptation (Weinberg, 2014)**

Type of Cellular Adaption	Definition
Anaplasia	Cells lack differentiation and specialised characteristics, often a feature of malignant tumours
Dysplasia	Disordered cell proliferation as result of changes in cellular arrangement, size and shape.
Neoplasia	A region of cellular growth that has escaped normal regulatory controls usually through the acquisition of DNA mutations. Solid neoplasms are commonly referred to as tumours
Hyperplasia	An area of increased and abnormal cell proliferation
Metaplasia	Cells that have transformed into another differentiated cell type

A more relevant question than what is cancer is perhaps to ask what are the characteristics of transformed cells? Table 1-2 Summarises features of cellular adaption that are seen in tissues with transformed cells. In 2000 Hanahan and Weinberg (Hanahan and Weinberg, 2000) described a process by which cells transform to become malignant by the acquisition of certain traits. This landmark paper identified central neoplastic traits that were named as the six hallmarks of cancer and have since been updated to add additional hallmarks (Table 1-3)

### Chapter 1.3.1 What is cancer?

(Hanahan and Weinberg, 2011). Criticisms of these hallmarks are that some features described are required for tumourigenesis and not specific for carcinogenesis (Lazebnik, 2010). However, it is important to note is that 'hallmarks' are defined in the paper as 'distinctive and complementary capabilities that enable tumour growth and metastatic dissemination' (Hanahan and Weinberg, 2000). An additional criticism is that the paper follows the somatic mutation theory (SMT) (Boveri, 2008) where cancer arises clonally from a single cell rather than cancer as a disease of disordered tissue organisation as proposed by the tissue organisation field theory (TOFT) (Sonnenschein and Soto, 2013). TOFT considers the altered biophysical and biomechanical properties of cancer tissue, something often overlooked in cancer biology. This landmark paper has been pivotal for summarising key features and properties of malignant cells but contains no reference to cancer causation.

**Table 1-3 Hanahan and Weinberg Hallmarks of Cancer**

Hanahan and Weinberg Hallmark	Description
Sustained proliferative signalling	Do not require external growth factor signals to divide
Evading growth suppressors	Obtain activating mutations capable of sustained downstream signalling
Avoiding immune destruction*	Evasion and inhibitor strategies to avoid tumour killing immune cells
Enabling replicative immortality	Cells can enter a non-proliferative senescent state. Rarely cancer cells escape from senescence with the ability to divide indefinitely (immortalisation).
Tumour-promoting inflammation*	Most tumours are infiltrated by immune cells. Inflammation can supply tumorigenic molecules to the tumour microenvironment. Chronic inflammation has been associated with carcinogenesis.
Activating invasion & metastasis	Acquire traits to invade surrounding healthy tissue and ability to metastasise or transit via the circulatory system to distant tissues
Inducing angiogenesis	Required to sustain nutrient availability and exclude metabolic waste in the area of increased cell mass
Genome instability & mutation*	DNA damage can result in mutations that may confer a selective advantage.
Resisting cell death	Evolve strategies to become resistant to programmed cell death (apoptosis).
Deregulating cellular energetics*	Altered cell metabolism is required for sustained cell growth and proliferation (Warburg, 1956).

\* = Added or updated in Hanahan & Weinberg 2011

### **1.3.2 Cell Cycle and DNA damage**

The eukaryotic cell cycle describes a methodical series of events including regulatory checks that are required in order for a cell to divide with high fidelity. There are four major phases of the cell cycle; G<sub>1</sub>, S, G<sub>2</sub> and M phase. During G<sub>1</sub> cells grow in size and synthesise the required components for cell division (Lodish *et al.*, 2013). In S phase DNA synthesis and replication happens, followed by another growth phase (G<sub>2</sub>). The final phase is M phase, or mitosis, whereby the cell divides to produce two identical daughter cells. Cells can also exit the cell cycle during G<sub>1</sub> to enter a quiescent, resting state known as G<sub>0</sub>.

Control of the cell cycle is governed by regulatory proteins, mainly cyclins and cyclin-dependent kinases (CDKs), at specific checkpoints throughout the cell cycle (Satyanarayana and Kaldis, 2009). Cyclins are specific to a cell cycle phase and can form complexes with CDKs to activate or inhibit regulatory proteins and control progression through the cell cycle.

Cells contain machinery to sense DNA lesions, amplify the damage signal, and initiate a cellular response, such as cell cycle arrest or beginning DNA repair, in a process known as the DNA damage response (DDR) (Jackson and Bartek, 2010). Different types of DNA damage can occur such as double stranded breaks (DSB) and DNA mismatches that induce specific DNA repair mechanisms. Identification of DNA damage and efficient repair are required by the cell to prevent the accumulation of mutations. A feature of cancer is that these cell cycle protective mechanisms, including the DDR, are dysregulated allowing for a cell to accrue mutations, avoid apoptosis and continue to proliferate (Ishikawa *et al.*, 2006). Incomplete DNA repair can have the consequence of increased mutagenesis and genomic instability, a hallmark of cancer (section 1.4.3).

The identification of DNA damage or cellular abnormalities at cell cycle checkpoints can result in various cellular outcomes. Initially cell-cycle arrest prevents proliferation of an abnormal cell, followed by induction of autophagy (a form of cellular degradation and recycling) and, if required, DNA repair. If the cell is severely damaged it will undergo programmed cell death otherwise known as apoptosis or enter a state of irreversible senescence where the cell is unable to proliferate.

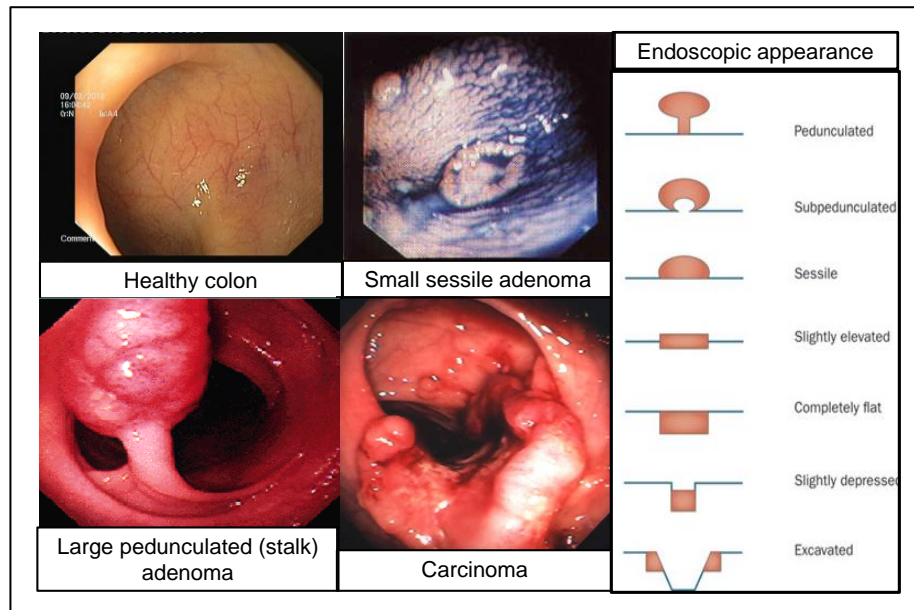


## 1.4 Colorectal Cancer (CRC)

### 1.4.1 Adenomatous polyps and colorectal cancer

The aetiological factors of CRC include genetics, diet and environmental factors however it is also a heterogeneous disease (Linnekamp *et al.*, 2015). Colorectal cancer is known to develop over many years and undergo a multi-step process. Initially hyperproliferation of the intestinal epithelium is seen leading to the development of an adenomatous polyp which may become cancerous defined by invasion of the tumour into the bowel wall (Vogelstein *et al.*, 2013). Not all polyps will become malignant and there currently exists a lack of predictive biomarkers (Lech *et al.*, 2016). Despite advances in the molecular analysis of CRC, methods of staging currently remain using clinicopathological criteria. The current clinical CRC prognosis criteria is the TNM staging system which takes into account Tumour size, inclusion of lymph Nodes and the presence of Metastatic cells (Amin *et al.*, 2017). An accumulation of evidence has led to the consensus that polyps may evolve into cancer over time via a locus of malignancy that develops within the tumour. One example of such evidence is that the removal of adenomas (polypectomy) during screening colonoscopies has been shown to reduce the incidence of CRC development (Winawer *et al.*, 1993).

Polyps can differ in their endoscopic and histological appearance, size and location within the colon (Hazewinkel and Dekker, 2011). Figure 1-4 shows images of colonic pathology taken during colonoscopy and the endoscopic appearance of various tumour types. Mucosal biopsies of newly diagnosed polyps are sent for histological analysis for sub-typing and information on malignant potential. Table 1-4 summarises key phenotypic and epidemiological features of different adenoma types based on the WHO Classification of Tumours of the Digestive System (Bosman *et al.*, 2010).



**Figure 1-4 Colonic tumourigenesis (adapted from Hazewinkel and Dekker 2011)**

Endoscopic images of a healthy colon, small sessile adenoma highlighted by indigo carmine dye spray, a large pedunculated adenoma and a carcinoma. The endoscopic appearance box highlights the different shapes of tumours seen in the colon and their terminology.

**Table 1-4 Summary of adenomatous polyp subtypes (Bosman *et al.*, 2010)**

Polyp Type	Neoplastic	Histological appearance	Endoscopic appearance	Estimated Prevalence
<b>Tubular</b>	Yes	Branched, irregular crypts in a fibrovascular stroma. Neutrophil infiltrate. Hyperchromatic nuclei	Pedunculated	65-80%
<b>Villous</b>	Yes	Long finger-like projections. More likely to have severe dysplasia	Sessile	5-10%
<b>Tubulovillous</b>	Yes	Features of tubular and villous types (see above)		10-25%
<b>Serrated</b>	Yes	Three subtypes: hyperplastic (HP). Sessile serrated adenoma (SSA), traditional serrated adenoma (TSA). For more detail see (Rosty and Bettington, 2014). Saw tooth appearance. High malignant potential	Sessile, often right sided	More common in women
<b>Inflammatory (pseudopolyps)</b>	No	Distorted crypt architecture, lamina propria immune cell infiltrate. Epithelial inflammation common		Mainly in IBD patients

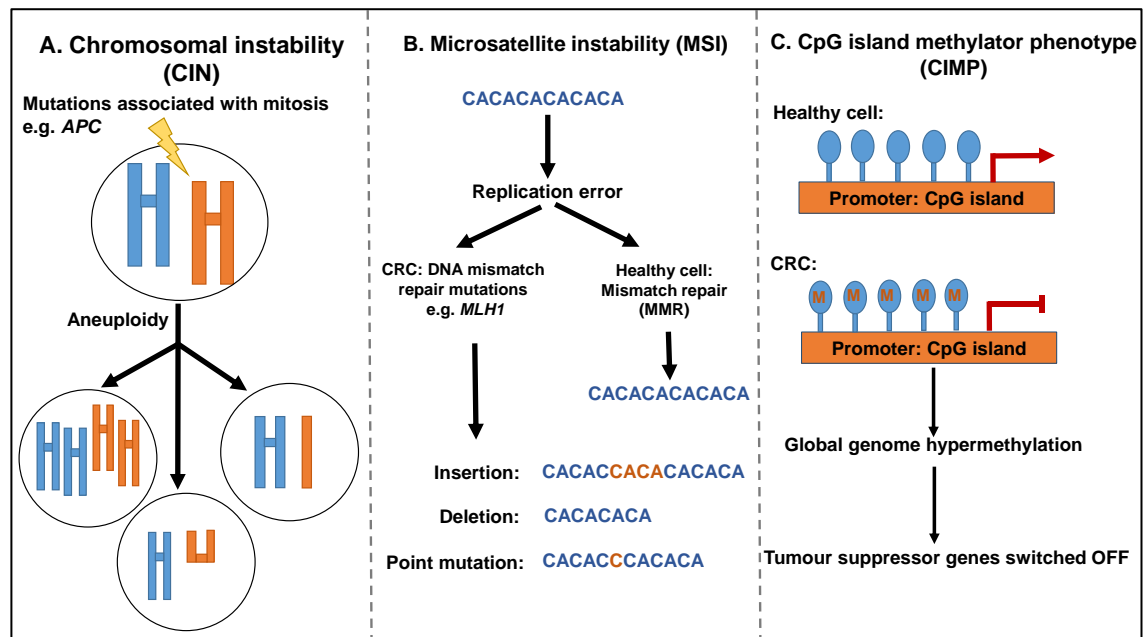
### **1.4.2 Genetic pathways of CRC**

There is a low rate of mutation in healthy human cells while large doses of carcinogens are likely to lead to cell death. For these reasons and from the study of specific cancer types it is thought that key events must occur on their own and in a sequential order for a cancer to form, often referred to as a genetic pathway (Knowles and Selby, 2005). The classic multi stage description of CRC carcinogenesis, sometimes referred to as the adenoma-carcinoma sequence or the 'Vogelgram', was described by Fearon and Vogelstein in 1988 (Vogelstein *et al.*, 1988). This model refers to the sequential progression of CRC which is accompanied by the temporal accumulation of somatic mutations in key genes/pathways, including the common mutations at specific steps of cancer progression. Since Fearon and Vogelstein the genomic heterogeneity of colorectal adenomas and carcinomas has been realised with additional pathways of CRC progression also described.

### **1.4.3 Genomic instability**

Genomic instability refers to chromosome aberrations and is a common feature of CRC. Two main types have been described, chromosomal instability (CIN) and microsatellite instability (MSI) (Figure 1-5). MSI occurs due to faulty DNA replication during S phase whereas CIN occurs due to a chromosome segregation error during mitosis. It is also possible for cells to acquire more than one type of genomic instability (Pikor *et al.*, 2013).

Chromosomal instability (CIN) refers to defective chromosome segregation with the outcome being an altered number of chromosomes (aneuploidy) or chromosome instability. The cause of CIN is linked to mutations in genes associated with mitosis, for example mutations resulting in centrosome mis-segregation (Rao and Yamada, 2013). A mutation in the adenomatous polyposis coli (APC) gene (section 1.4.4) is common in the early stages of adenoma development and is linked to CIN, whether mutated APC causes CIN or vice versa is of some debate (Rusan and Peifer, 2008). This type of genomic instability is also associated with poor prognosis in CRC (Hveem *et al.*, 2014).



**Figure 1-5 Types of Genomic Instability in Colorectal Cancer**

(A) Chromosomal instability refers to alterations in chromosome number or structure often resulting from mutations associated with mitosis.

(B) Microsatellite repeat sequences can result in replication errors such as an insertion, deletion or point mutation. Mutations in DNA mismatch repair in CRC do enable repair of these errors.

(C) Hypermethylated CpG promoters in colorectal cancer can result in tumour suppressor genes being switched off. CpG sites are represented by blue 'lollipops'. 'M' denotes a methylated CpG site.

Microsatellites are tandem repeats of short DNA sequences (1-5 bp) that can be prone to errors during replication from slippage of the DNA polymerase. In a healthy cell these microsatellite alterations are compensated for by the existence of mismatch repair (MMR) mechanisms. The primary contributor in CRC to microsatellite expansions and contractions, in other words MSI, (Boland and Goel, 2010) (Figure 1-5B) are mutations in important MMR genes such as *MLH1*. Germline mutations in MMR genes result in a syndrome known as hereditary nonpolyposis colorectal cancer (HNPCC) or Lynch syndrome whereby these individuals have an increased risk of developing CRC (Lynch and Krush, 1971, Douglas *et al.*, 2005).

An alternative molecular phenotype of CRC development an epigenetic instability pathway known as the CpG island methylator phenotype (CIMP) was first described by Toyota *et al.*, 1999 (Figure 1-5C). Some gene promoters are rich in CG dinucleotides known as CpG islands. A common form of epigenetic regulation is via methylation of CpG islands and consequential

transcriptional silencing. CIMP describes a subset of CRC with hypermethylated promoters that result in switched off tumour suppressor genes (Nazemalhosseini Mojarad *et al.*, 2013).

#### **1.4.4 Importance of APC and $\beta$ -catenin signalling in CRC**

Wnt signalling is a significant pathway for embryonic development, maintenance of cellular homeostasis and wound healing (Clevers, 2006). These important cellular functions are mediated through fine regulation of the abundance and localisation of the transcriptional co-activator,  $\beta$ -catenin. When located within the nucleus  $\beta$ -catenin interacts with the TCF/LEF transcription factor to induce gene expression including the protooncogenes *c-Myc* and *cyclin D1*. Regulation of total  $\beta$ -catenin is mediated by the  $\beta$ -catenin destruction complex, which comprises of the proteins, APC, Axin, glycogen synthase kinase 3 (GSK-3) and casein kinase 1 (CK1) (MacDonald *et al.*, 2009). In the canonical Wnt signalling pathway the destruction complex can be activated by Wnt ligand binding to an upstream membrane-bound receptor. In the absence of Wnt activation  $\beta$ -catenin is phosphorylated and targeted for degradation to the proteasome. Wnt ligand activation promotes  $\beta$ -catenin separation from the cytoplasmic protein complex and subsequent translocation to the nucleus thereby inducing gene expression.

Cross-talk with other pathways such as Hippo signalling (Konsavage Jr *et al.*, 2012, Kim and Jho, 2014) enables fine regulation of  $\beta$ -catenin quantity and location. The main pool of  $\beta$ -catenin is associated with cadherin-mediated cell adhesion.  $\beta$ -catenin can be sequestered at the intercellular adherens junctions via its binding to E-cadherin. Cleavage of E-cadherin results in  $\beta$ -catenin dissociation and accumulation in the nucleus.

Evidence for the importance of  $\beta$ -catenin signalling in CRC is observed based on the fact that 80% of CRC tumours have acquired mutations within this pathway (Rowan *et al.*, 2000). APC is a potent CRC tumour suppressor with somatic mutations resulting in loss of function or production of a truncated protein (Rowan *et al.*, 2000). Germline mutations in APC are linked to the hereditary condition familial adenomatous polyposis (FAP) whereby individuals develop multiple (over a hundred) colonic adenomatous polyps (Half *et al.*, 2009). A common mouse model of intestinal tumourigenesis is the minimal intestinal neoplasia mouse model (*Apc*<sup>(Min/+)</sup>). These mice are heterozygous for a point mutation in the APC gene resulting in the development of multiple intestinal adenomas (Su *et al.*, 1992).

*APC* mutations in CRC often occur within a portion of the gene referred to as the mutation cluster region (MCR), a region required for  $\beta$ -catenin binding (Nagase and Nakamura, 1993). Aside from its role as a scaffold protein within the  $\beta$ -catenin destruction complex *APC* also has other functions that may be of relevance to its association with CRC. *APC* can bind to and stabilise microtubules (Mimori-Kiyosue *et al.*, 2000, Zumbrunn *et al.*, 2001) thereby associating *APC* with cell division and CIN. Not all adenomas with mutant *APC* are aneuploidy suggesting that mutations in *APC* predate and therefore could cause CIN (Sieber *et al.*, 2002). *APC* can also bind to components of the DNA repair mechanism base-excision repair linking it with cell fate of DNA damaged cells and MSI (Jaiswal and Narayan, 2008).

#### **1.4.5 Colonoscopy**

Colonoscopy is an endoscopic procedure considered the 'gold standard' for the detection of gastrointestinal pathology such as colonic polyps or inflammation. Clinical indications for referral include anaemia, unexplained changes in bowel habit, unexplained weight loss and rectal bleeding. Following bowel preparation medication, a colonoscope is passed through the length of the colon through the ileocaecal valve and ending at the terminal ileum. The colonoscope is then removed gradually with patient position changes to look carefully for any tumours. Suspected regions of abnormal mucosa can be stained using a dye such as indigo carmine dye (Figure 1-4). Despite this there is still an estimated 2% polyp miss rate (Hazewinkel and Dekker, 2011). On finding a polyp/neoplasm some biopsies are taken for histological assessment and the polyp itself removed by a procedure known as a polypectomy.

### **1.5 Bacterial pathogenesis and virulence factors**

The red queen hypothesis of evolution describes an evolutionary arms race between competing species for finite resources i.e. 'it takes all the running you can do to stay in the same place' (Van Valen, 1973). An updated version of this evolutionary hypothesis details how species expand in population size to a peak, with a failure to expand resulting in a decline towards extinction (Zliobaite *et al.*, 2017). Competing organisms therefore require genome plasticity and the acquisition of mutations to adapt and evolve against competitors. Within a single bacterial species there can exist a large genomic diversity between individual isolates or strains and this

### Chapter 1.5.1 Bacterial toxins

has been highlighted with the advancement of DNA sequencing technology and subsequent surge in comparative genomics. The pan-genome of a species comprises a core genome present in all isolates and an accessory genome present in only some strains. As more genomes of a species are sequenced the number of unique genes identified decreases however extrapolating the curve suggests that a non-zero asymptotic value is reached (Tettelin *et al.*, 2005).

Bacterial pathogens carry virulence factors that play a role in disease pathogenesis (Brogden *et al.*, 2000). Loss of virulence factors generally does not affect viability of the organism but does affect virulence. Commonly these virulence factors are associated with colonisation (e.g. adherence factors), evasion of the host immune response (e.g. capsule formation), host tissue invasion or translocation of effector molecules into host cells via secretion systems or damage to the host via toxin production. Virulence factors are often part of the aforementioned accessory genome coded for on DNA elements such as prophages, transposons, plasmids and pathogenicity islands (Brown *et al.*, 2012). This suggests these genes have either been acquired recently in evolutionary terms or are easily transmissible between organisms.

An interesting evolutionary question regarding pathogenicity is what benefit does the pathogen gain by damaging or even killing its host? The answer to this is likely to be subtly different on a case-by-case basis but an interesting example is *Clostridium botulinum*, which has a very potent virulence factor in botulinum toxin. This toxin is easily able to kill the human host but the bacterium itself is soil-based and cannot grow/thrive in humans therefore humans are an ecological sink for the bacterium (Sokurenko *et al.*, 2006). A comprehensive review on the topic of virulence factors and their evolution is described by (Brown *et al.*, 2012).

#### **1.5.1 Bacterial toxins**

Toxins are common bacterial virulence factor and effectors of bacterial pathogenesis (Rudkin *et al.*, 2017). Traditionally bacterial toxins were defined as causing cellular damage and death but it has become apparent that they can also attack and hijack host cell function. Various subtypes of bacterial toxins exist. Pore-forming toxins can induce pore formation in cellular membranes, intracellular toxins act within the cell and the third category mimic host signalling molecules.

### **1.5.2 *B. fragilis* – an example of a pathobiont**

Opportunistic pathogens or pathobionts usually live in symbiosis with the host (commensals) but under certain conditions can become pathogenic. Often opportunistic pathogens are compartmentalised into environments where they cause disease and locations where they are commensal (Brown *et al.*, 2012). For example, *B. fragilis* can be commensal within the healthy GI tract but following disruption of the intestinal wall may trigger intra-abdominal abscess formation and sepsis (Wexler, 2007).

#### **1.5.2.1 BfUbb**

Whole genome sequencing of the *B. fragilis* type strain NCTC9343 revealed a low GC% region, suggesting the presence of a pathogenicity island which was found to include a protein with homology to human ubiquitin (Patrick *et al.*, 2011). Expression of BfUbb has been confirmed and the protein is likely to be delivered to cells in outer membrane vesicles (OMVs). BfUbb differs from human ubiquitin due to a C-terminal cysteine instead of a glycine residue. This cysteine residue can form a disulphide bridge with human E1 activating enzyme. BfUbb inactivates E1 and is likely to have profound effects on the ubiquitination cascade and subsequently the host cell, and therefore is a putative virulence factor (Patrick and Blakely, 2012).

Although BfUbb is the first bacterial protein identified with direct homology to ubiquitin other bacterial pathogens have been shown to modulate the host ubiquitination system using a variety of mechanisms (Zhou and Zhu, 2015).

## **1.6 Microbial-induced cancer**

A range of epidemiological studies have reported associations between bacterial pathogens and/or microbiome dysbiosis and an increased incidence of a specific cancer. Table 1-5 lists examples of bacteria and bacteria-associated diseases that have been epidemiologically linked to cancer.



**Table 1-5 Bacterial pathogens and bacterial dysbiosis epidemiologically linked to a specific cancer type**

Authors	Bacterium or bacterial dysbiosis	Cancer type	Risk
(Michaud <i>et al.</i> , 2013)	<i>Porphyromonas gingivalis</i>	Pancreatic	Elevated <i>P. gingivalis</i> antibody titre associated with a two-fold increased risk
(Abnet <i>et al.</i> , 2001)	Poor oral health as indicated by tooth loss	Oesophageal and gastric	Oesophageal RR: 1.3 Gastric non-cardia RR: 1.8
(Zeng <i>et al.</i> , 2016) Meta-analysis	Periodontal disease	Lung	HR = 1.24 adjusted for smoking status
(Randi <i>et al.</i> , 2006)	<i>S. typhi</i> and <i>S. paratyphi</i>	Gallbladder	Pooled RR: 4.8

RR = Relative risk; HR = Hazard ratio

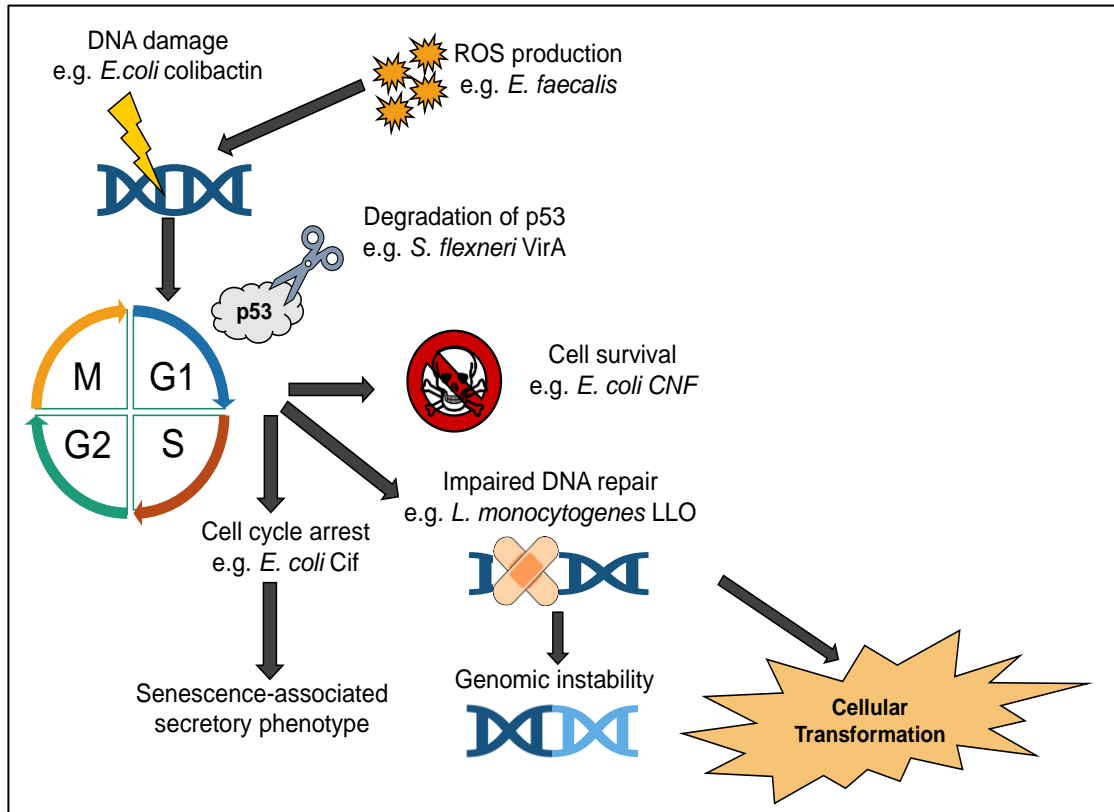
The aforementioned examples are associations and therefore not necessarily causative. Potential oncogenic mechanisms of specific bacteria and their virulence factors have been identified and are detailed in section 1.6.2.

### 1.6.1 HPV - a viral example

Arguably the best evidence for causality of cancer by a microorganism is available for Human Papilloma Virus (HPV). In the UK vaccinations to prevent colonisation of HPV are now routinely given to teenage girls to reduce the incidence of cervical cancer (Drolet *et al.*, 2015). The primary mechanism of tumorigenesis is executed by the HPV oncoproteins (E6 and E7) which can inactivate p53 and retinoblastoma (pRB) tumour suppressor proteins thereby driving aberrant cell proliferation (DeFilippis *et al.*, 2003). The well-characterised oncogenic mechanisms of HPV provide insight into microbe-induced tumourigenesis. Microbe-driven nuclear localisation of  $\beta$ -catenin has also been observed as an oncogenic characteristic of HPV (Rampias *et al.*, 2010).

### 1.6.2 Bacterial mechanisms of cell cycle modulation

Putative carcinogenic bacteria have a range of mechanisms by which they can induce cellular transformation. Impairment of cell cycle regulation and DNA damage response (Figure 1-6), modulation of host cell signalling (Figure 1-7) and direct immune system perturbation are the main categories by which bacteria can push cells towards a malignant phenotype. Comprehensive reviews on this subject can be found in (Lax, 2005) and (Gagnaire *et al.*, 2017).



**Figure 1-6 Bacterial impairment of cell cycle regulation and DNA damage response**

Examples of mechanisms used by bacterial pathogens to disrupt cell cycle regulation and the DNA damage response subsequently leading to cellular transformation. ROS = Reactive Oxygen Species

Bacteria can directly damage DNA by a number of mechanisms. These include the production of the genotoxin colibactin in *E. coli* (section 1.7.2) that induces double strand breaks (Nougayrede *et al.*, 2006). Host cell cycle checkpoints exist to identify and initiate repair of DNA damage although repair may be incomplete while chronic DNA damage can result in genomic instability or cellular senescence. Another bacterial genotoxin, cytolethal distending toxin (CDT), when chronically applied to cells at a sublethal dose, blocks initiation of DDR and cells transform towards a malignant phenotype (Guidi *et al.*, 2013). Bacteria can also damage DNA indirectly; for example *Enterococcus faecalis* can produce extracellular superoxide, a precursor of many types of reactive oxygen species (ROS) (Wang *et al.*, 2012). DNA damage and genomic instability as a result of *E. faecalis* superoxide production has been shown *in vitro* and *in vivo* (Huycke and Moore, 2002).

### Chapter 1.6.3 Oncogenic modulation of host cell signalling by bacteria

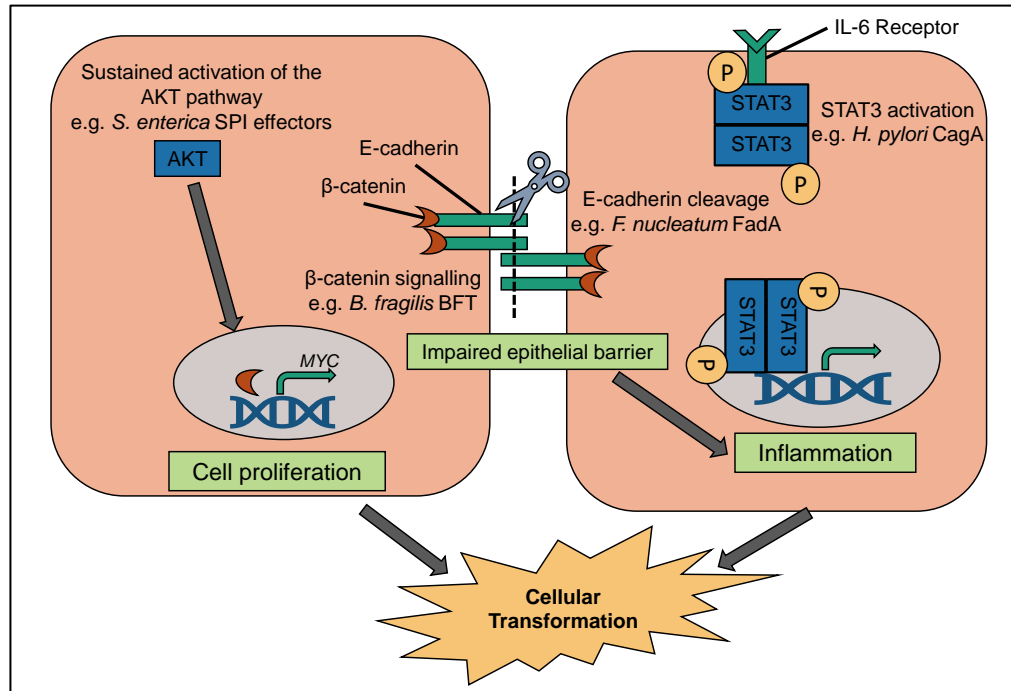
Cycle-inhibiting factor (Cif) was first discovered in *E. coli* in 1997 (De Rycke *et al.*, 1997) when it was observed that it induced actin stress fibres and enlarged cells. Cif can also be found in other bacteria: *Yersinia pseudotuberculosis*, *Photobabdu luminescens*, *Photobabdu asymbiotica* and *Burkholderia pseudomallei* (Jubelin *et al.*, 2009). Cif does not trigger DNA damage but instead interacts with host protein NEDD8 resulting in inhibited ubiquitin ligase activity (Cui *et al.*, 2010, Jubelin *et al.*, 2010, Morikawa *et al.*, 2010), accumulation of p21<sup>Waf1/cip1</sup> and p27<sup>kip2</sup> (Samba-Louaka *et al.*, 2008) and inhibitory phosphorylation of cyclin dependent kinase-1 (CDK-1). The consequence is cell cycle arrest in the host cell.

Other bacterial mechanisms do not damage DNA but impair the DNA damage response. For example, *Listeria monocytogenes* toxin listeriolysin O (LLO) induces DNA damage and triggers degradation of the DDR protein Mre11, a mechanism required for sustained *L. monocytogenes* infection (Samba-Louaka *et al.*, 2014).

*E. coli* Cytotoxin necrotizing factor CNF1 promotes cell survival via the Akt/I $\kappa$ B pathway, ultimately leading to Bcl-2 overexpression and prevention of apoptosis (Miraglia *et al.*, 2007). The *Salmonella flexneri* effector VirA can also inhibit apoptosis via the degradation of p53 (Bergounioux *et al.*, 2012). VirA achieves this by activating the host protease calpain to degrade p53. Other bacteria have been found to manipulate the p53 pathway (Zaika *et al.*, 2015), including *Helicobacter pylori* cytotoxin-associated gene-A (CagA) (Buti *et al.*, 2011) and *Chlamydia trachomatis* (Siegl *et al.*, 2014).

### 1.6.3 Oncogenic modulation of host cell signalling by bacteria

Cellular transformation can be promoted by bacterial modulation of host cell signalling pathways (Figure 1-7). Many of the mechanisms result in sustained cell proliferation which may increase the rate of cell transformation (Lax and Thomas, 2002). One mechanism used by bacteria to initiate cell proliferation is to activate  $\beta$ -catenin signalling (section 1.6.5). This can be achieved either through disruption of adherens junctions either through direct E-cadherin cleavage, for example by the *Fusobacterium nucleatum* adhesin FadA (Rubinstein *et al.*, 2013) or indirectly via  $\gamma$ -secretase as occurs in *B. fragilis* toxin (BFT) (Wu *et al.*, 2007). The other consequence of targeting adherens junctions in epithelial cells is that this impairs the barrier property of the epithelium allowing for paracellular migration of the microbiota and consequent inflammation.



**Figure 1-7 Bacterial modulation of host cell signalling**

Mechanisms used by bacterial pathogens to modulate host cell signalling, driving the cell towards cellular transformation.

*S. enterica* SPI effectors can induce sustained activation of the AKT and MAPK pathways initiating a range of cell signalling that can lead to cell proliferation, growth and differentiation and ultimately cellular transformation (Scanu *et al.*, 2015). Another example of an oncogenic pathway targeted by bacterial pathogens is Signal transducers and activators of transcription 3 (STAT3). *H. pylori* infection results in phosphorylation and therefore activation of (STAT3) in a CagA mediated process (Bronte-Tinkew *et al.*, 2009). When activated, dimeric STAT3 can translocate to the nucleus to act as a transcription factor and in epithelial cells will initiate cell proliferation, migration, and transformation. In immune cells STAT3 activation results in a pro-inflammatory phenotype (Cheng *et al.*, 2003).

#### 1.6.4 *Helicobacter pylori* - a bacterial example

The most established link between bacteria and cancer causation is between *H. pylori* and gastric adenocarcinoma. Originally the connection between *H. pylori* and cancer was unclear but a convincing factor was the evidence that eradication of the bacterium was associated with a significant decrease in the development of premalignant lesions (Wong *et al.*, 2004, Mera *et*

#### Chapter 1.6.4 *Helicobacter pylori* - a bacterial example

*al.*, 2005). *H. pylori* can cause inflammation of the lining of the stomach (gastritis) (Marshall *et al.*, 1985) and if left untreated this long-term inflammation has been associated with malignancy (Uemura *et al.*, 2001). Extensive literature exists on mechanisms of *H. pylori*-induced tumourigenesis and therefore provides an excellent model when investigating other links between bacteria and cancer. Numerous *H. pylori* virulence factors have been attributed to the carcinogenic process but primarily the toxins, CagA and Vacuolating cytotoxin A (VacA), are considered instrumental.

The *cagA* gene is located on the *cag* pathogenicity island (PAI) that codes for a type IV secretion system and is not present in all *H. pylori* strains. CagA+ strains have been shown to increase the risk of *H. pylori*-induced gastric cancer (Parsonnet *et al.*, 1991). The *cag* type IV secretion system (T4SS) is a molecular syringe that allows for the injection of the CagA effector protein into a host cell (Odenbreit *et al.*, 2000). Inside the cell CagA becomes phosphorylated at the amino acid motif EPIYA by host Src kinases (Stein *et al.*, 2002). CagA then binds and deregulates SRC homology 2 domain (SH2)-containing tyrosine phosphatase SHP-2 (Higashi *et al.*, 2002b) to initiate a range of cell signalling events that push the cell towards malignancy. *H. pylori* strains can differ in their number of EPIYA motifs altering binding affinity for the host SHP-2 phosphatase and therefore have been associated with the carcinogenic potential of a strain (Higashi *et al.*, 2002a). T4SS expression and CagA delivery into the cell occurs basolaterally (Tegtmeyer *et al.*, 2017). To permit paracellular transmigration disruption of the tight and adherens junctions is carried out by another *H. pylori* virulence factor, the serine protease high-temperature requirement A (HtrA) (Hoy *et al.*, 2010).

The VacA toxin is multifunctional but named after its ability to bind cells, become internalised and induce cytoplasmic vacuolation of host cells (Leunk *et al.*, 1988). Following internalization into endosomal-like vesicles (Papini *et al.*, 1994) VacA forms anion-selective channels (Iwamoto *et al.*, 1999) in the vesicle membrane, triggering osmotic swelling (Tombola *et al.*, 1999) and vacuolation. VacA can indirectly activate multiple cellular signalling pathways including, p38 and ERK1/2 (Nakayama *et al.*, 2004) and has also been shown to cause damage to the mitochondria (Kimura *et al.*, 1999) triggering apoptosis (Galmiche *et al.*, 2000). All strains contain the *vacA* gene but there are variations in sequence with some genotypes linked with an increased risk of developing gastric cancer (Miehlke *et al.*, 2000). The ability of VacA to target

#### Chapter 1.6.5 Modulation of $\beta$ -catenin signalling

immune cells provides an effective method of immune evasion which facilitates the persistence of *H. pylori* (Djekic and Muller, 2016).

Tumour necrosis factor- $\alpha$  (TNF- $\alpha$ ) inducing protein (Tip $\alpha$ ) is an additional putative carcinogenic toxin produced by *H. pylori* (Suganuma *et al.*, 2008). Tip $\alpha$  induces NF $\kappa$ B activation resulting in increased chemokine and TNF- $\alpha$  secretion.

Other *H. pylori* virulence factors that have been proposed to be associated with gastric tumourigenesis are Outer inflammatory protein A (OipA) (Yamaoka *et al.*, 2000) and Duodenal ulcer promoting gene a (*dupA*) (Lu *et al.*, 2005a). The OipA gene can be present in strains in functional (on) or non-functional (off) forms with a meta-analysis concluding that the on form is significantly associated with gastric cancer (Liu *et al.*, 2013). Putative oncogenic mechanisms of OipA include aiding translocation of CagA (Horridge *et al.*, 2017), activating p38 and NF $\kappa$ B pathways to induce interleukin-6 (IL-6) secretion (Lu *et al.*, 2005b) and suppressing apoptosis in gastric cells (Al-Maleki *et al.*, 2017).

The *dupA* gene is strongly associated with duodenal ulcer formation but preventative against gastric cancer (OR = 0.41, (Lu *et al.*, 2005a). *dupA* and the surrounding gene cluster is proposed to form a T4SS with DupA being an effector protein (Jung *et al.*, 2012). DupA exhibits ATPase activity, is associated with increased IL-8 production and aids *H. pylori* tolerance to the low stomach pH via increased urease secretion (Wang *et al.*, 2015). A possible anti-cancer property of DupA is its ability to trigger apoptosis (Wang *et al.*, 2015). Conflicting results exist in the literature as to the relationship between *dupA* and gastric cancer. A meta-analysis found a positive association between *dupA* and duodenal ulcer but no significant association with gastric cancer (Shiota *et al.*, 2010).

#### **1.6.5 Modulation of $\beta$ -catenin signalling**

The finely regulated Wnt/ $\beta$ -catenin signalling pathway (previously introduced in section 1.4.4) is targeted by many bacterial pathogens in diverse ways. As many genes transcriptionally activated by  $\beta$ -catenin are involved in cell proliferation and differentiation this pathway is often exploited in microbial-induced cancer.

#### Chapter 1.6.5 Modulation of $\beta$ -catenin signalling

*H. pylori* targets many components of  $\beta$ -catenin signalling using both species and strain-specific virulence factors. A common mechanism of bacterial modulation of  $\beta$ -catenin is to target the E-cadherin associated pool of  $\beta$ -catenin either through binding to or cleaving E-cadherin. *H. pylori* can influence E-cadherin directly and indirectly to stimulate nuclear accumulation of  $\beta$ -catenin. *H. pylori* Toll-like Receptor 2 (TLR2) signalling can induce E-cadherin cleavage via the host protease, calpain (O'Connor *et al.*, 2011) while the gastric cancer associated protein CagA has been shown to directly bind to E-cadherin with resultant upregulated gene expression (Murata-Kamiya *et al.*, 2007). *H. pylori* vacuolating cytotoxin A (VacA) targets a different aspect of  $\beta$ -catenin signalling by activating the epidermal growth factor receptor (EGFR) with downstream inactivation of the destruction complex protein GSK3 $\beta$  (Nakayama *et al.*, 2009).

The *F. nucleatum* FadA adhesin and BFT also use mechanisms of E-cadherin/ $\beta$ -catenin mediated oncogenesis (Wu *et al.*, 2003, Rubinstein *et al.*, 2013). These organisms and virulence factors are described in more detail in sections 1.7.4 and 1.7.5 respectively.

Previously considered non-pathogenic the *Salmonella* typhimurium strain PhoP<sup>c</sup> decreases levels of ubiquitinated  $\beta$ -catenin and promotes its nuclear accumulation. This strain was shown to increase c-Myc expression and cell proliferation via the effector protein AvrA (Sun *et al.*, 2004). These findings were replicated in the AOM/DSS mouse model of intestinal neoplasia with the consequence of altered  $\beta$ -catenin signalling being that AvrA+ *Salmonella* strains significantly increased tumour incidence compared with AvrA- strains and no bacteria (Lu *et al.*, 2014).

An example of an oral pathogen with a link to  $\beta$ -catenin and carcinogenesis is *Porphyromonas gingivalis*. The *P. gingivalis* proteases gingipains can proteolytically cleave  $\beta$ -catenin and members of the destruction complex (Zhou *et al.*, 2015). These  $\beta$ -catenin cleavage products can translocate to the nucleus where it is functionally active. *P. gingivalis* treatment has been shown to accelerate the cell cycle and invasive abilities of human immortalized oral epithelial cells (HIEOECs) (Katz *et al.*, 2000). Taken together these results suggest *P. gingivalis* infection may transform oral epithelial cells and promote carcinogenesis.

The pattern recognition receptor toll-like receptor 4 (TLR4) is overexpressed in colitis-associated and sporadic CRC (Santaolalla *et al.*, 2013). TLR4 is an innate immune receptor that

### Chapter 1.7.1 Hypotheses of microbiota-induced CRC

recognises bacterial lipopolysaccharide (LPS). TLR4 can activate the  $\beta$ -catenin pathway therefore TLR4 overexpression can lead to increased crypt cell proliferation. Whilst not an example of a specific bacterial pathogen modulating  $\beta$ -catenin signalling it highlights the importance between bacteria,  $\beta$ -catenin and CRC (Santaolalla *et al.*, 2013).

Viral pathogens, for example Rift Valley fever virus (RVFV), can also modulate  $\beta$ -catenin signalling, possibly to manipulate cell cycle conditions for viral replication (Harmon *et al.*, 2016).

## 1.7 CRC microbiome

### 1.7.1 Hypotheses of microbiota-induced CRC

A variety of hypotheses have been proposed by which the microbiome may induce/drive colorectal carcinogenesis. The alpha bug hypothesis (Sears and Pardoll, 2011) describes the role of the microbiome in the tripartite relationship between colonic epithelial cells, inflammation and microbiome in the development of CRC. The theory describes how 'alpha bugs' crowd out protective species and have pro-oncogenic characteristics that can trigger a cascade that results in colonic tumour growth as per the 'Volgelgram' model (section 1.4.2). The driver-passenger model for CRC builds on the concept of alpha-bugs instead referring to these as 'driver' organisms. A focus of this hypothesis is that the local microenvironment of the tumour site will change temporally as the healthy epithelium progresses into an adenoma and finally a carcinoma. A consequence of this environmental change is altered selection pressures on the associated microbiome. It is proposed that initial driver microbes have properties that can initiate tumourigenesis whereas 'passenger' species have a competitive advantage in this altered tumour milieu allowing for their colonisation. Passenger species may have either tumour promoting or suppressive characteristics (Tjalsma *et al.*, 2012).

A significant aspect of unpicking mechanisms of microbial driven CRC will be to make clear distinctions between the CRC-associated microbiome and CRC-causing (and CRC-preventing) microbes.



### 1.7.2 *Escherichia coli* toxins and CRC

Bacterial toxins linked to CRC (briefly introduced in section 1.6.3) can be divided into two groups; genotoxins due to their DNA-damaging properties and toxins that manipulate host cell signalling. Phylogenetic analysis divides *E. coli* strains into four main phylotypes (A, B1, B2 and D) (Selander *et al.*, 1986). Phylotypes B2 and D tend to contain strains with more virulence factors (Picard *et al.*, 1999), including a range of putative tumorigenic toxins.

Cytolethal distending toxin (CDT) is a holotoxin made up of three proteins, CdtA, CdtB and CdtC. CDT is not only found in *E. coli* but also in multiple other species of Gram-negative bacteria including *S. typhi* and *Campylobacter jejuni*. CdtB has homology to human DNase and indeed cells treated with CDT exhibit markers of DNA damage (Lara-Tejero and Galan, 2000). CDT can denature or relax supercoiled plasmid DNA *in vitro* while nuclear translocation of CdtB has also been observed (McSweeney and Dreyfus, 2004). Various DNA damage responses have been shown to be induced by cell exposure to CDT treatment (Fedor *et al.*, 2013, Fahrer *et al.*, 2014). Following the DDR in cells treated with CDT they enter irreversible G<sub>2</sub>/M cell cycle arrest (Poon *et al.*, 1997, Bielaszewska *et al.*, 2005).

The *pks+* genotoxicity island (Nougayrede *et al.*, 2006) encodes machinery for the synthesis of peptide-polyketide hybrid compounds, including synthesis of the genotoxin colibactin. Colibactin is another *E. coli* genotoxin able to induce double-stranded DNA breaks (Viljoen *et al.*, 2015, Vizcaino and Crawford, 2015). *E. coli* induced genotoxicity triggers cell cycle arrest and cell death (Nougayrede *et al.*, 2006). Cells treated with colibactin are enlarged (megacytosis) which is a sign of incomplete DNA repair and genomic instability (Cuevas-Ramos *et al.*, 2010). A common CRC mouse model when treated with colibactin-producing *E. coli* develop an increased number of colonic tumours in a mechanism dependent on the senescence-associated secretory phenotype (SASP) (Cognoux *et al.*, 2014).

An *E. coli* toxin that is not genotoxic is cytotoxic necrotizing factor (CNF). CNF deamidates a specific glutamine residue of the small GTPases RhoA, Rac and Cdc42 (Schmidt *et al.*, 1998) leading to cytoskeletal alterations (Fiorentini *et al.*, 1988) and senescence. CNF can prevent apoptosis by a mechanism previously mentioned in section 1.6.3. Cif also does not induce genotoxicity but similarly to CNF is a cyclomodulin thereby initiating cell cycle arrest.

The literature shows a lack of consensus regarding the association between *E. coli* and CRC in humans (Table 1-6). Geographical location and tumour site within the colon may be important factors to explain variation seen amongst the studies.

**Table 1-6 *E. coli* findings in human colorectal cancer (CRC) samples**

Author	Sample origin	Findings
(Buc <i>et al.</i> , 2013)	France	Increased prevalence of <i>cdt</i> , <i>cif</i> , <i>cnf</i> and <i>pks</i> in tumour samples compared with diverticulosis
(Bonnet <i>et al.</i> , 2014)	France	Tumours contained an increased level of mucosa-associated and internalised <i>E. coli</i>
(Raisch <i>et al.</i> , 2014)	France	<i>Pks</i> and <i>cnf</i> genes were significantly more prevalent in CRC than diverticulosis. CRC <i>E. coli</i> strains had low levels of adhesion and invasion but increased capacity to form biofilms
(Viljoen <i>et al.</i> , 2015)	South Africa	High level colonisation of <i>pks</i> -positive <i>E. coli</i> was significantly associated with tumour samples
(Gagniere <i>et al.</i> , 2017)	France	<i>Pks</i> status was associated with MSI CRC phenotype
(Shimpoh <i>et al.</i> , 2017)	Japan	No significant difference in <i>pks</i> status amongst control, adenoma and CRC patients.

### 1.7.3 *Streptococcus gallolyticus* subsp. *gallolyticus*

Historically the bacterium with the longest association to CRC is *Streptococcus gallolyticus*. Originally named *Streptococcus bovis* this and related taxa have subsequently been subdivided with biotype I (now classified as *Streptococcus gallolyticus* subsp. *gallolyticus*, (Schlegel *et al.*, 2003) having the strongest link with CRC bacteremia. The original association with CRC was due to the observation that 65-70% individuals with *S. bovis* endocarditis had previously undiagnosed gastrointestinal tumours (Bayliss *et al.*, 1984, Beeching *et al.*, 1985). A meta-analysis of all studies linking *S.bovis/S.gallolyticus* to CRC confirmed an increased prevalence in adenoma/carcinoma cases than healthy individuals. Interestingly, a higher prevalence (43%) was associated with adenomas compared with 18% for carcinomas indicative perhaps of an initiating/driver role in tumourigenesis (Boleij *et al.*, 2011b). Seropositivity to peptidoglycan anchored pilin proteins has also been shown to be significantly higher in CRC cases compared to controls (Butt *et al.*, 2016) (Butt *et al.*, 2016). Boleij *et al.*, 2011 discovered invasive strains that could translocate across polarised Caco-2 cells via a paracellular mechanism. Translocation of the bacteria may also explain seropositivity and link with bacteremia. *S. gallolyticus* may have a colonisation advantage, and therefore be a candidate passenger

#### Chapter 1.7.4 *Fusobacterium nucleatum*

bacterium, due to its ability to bind intestinal mucins, fibrinogen (Martins *et al.*, 2015, Martins *et al.*, 2016), and collagen rich tumour tissue (Boleij *et al.*, 2011a). A recent report, however, provides evidence of a *S. gallolyticus* tumour driver mechanism in that the bacterium can increase cell proliferation via  $\beta$ -catenin signalling in cultured cell lines (Kumar *et al.*, 2017). Evidence that this translates to an *in vivo* environment was shown in a mouse model of CRC as mice orally gavaged with *S. gallolyticus* developed a significantly higher tumour burden than mice challenged with control bacteria (Kumar *et al.*, 2017).

#### **1.7.4 *Fusobacterium nucleatum***

*Fusobacterium nucleatum* is a heterogeneous species currently classified into 5 subspecies; *animalis*, *fusiforme*, *vincentii*, *polymorphum* and *nucleatum* (Allen-Vercoe *et al.*, 2011). Recent publications suggest *fusiforme* and *vincentii* do not differ sufficiently to be classed as separate subspecies and indeed the remaining subspecies should be classified as individual species in their own right (Manson McGuire *et al.*, 2014, Kook *et al.*, 2017). These bacteria are commonly found in the oral cavity and are linked with odontogenic and periodontal disease. *F. nucleatum* is a highly adhesive species able to invade host cells (Manson McGuire *et al.*, 2014) and is also recognized as a 'bridging organism' in the development of dental plaque, a multi-species biofilm (Kaplan *et al.*, 2009). The mechanism of adhesion and invasion is by the formation of a complex between secreted and anchored forms of the FadA adhesin (Xu *et al.*, 2007). The complex binds to a host receptor, including the adherens junction protein E-cadherin (Rubinstein *et al.*, 2013), and is then able to facilitate invasion of the bacterium into the host cell following internalisation of FadA. Strains of *F. nucleatum* can vary in their invasive capacity as strains isolated from IBD patients were found to be more invasive than strains isolated from healthy individuals (Strauss *et al.*, 2011).

16S rRNA and metagenomic sequencing of the CRC microbiome has led to the discovery of specific bacterial taxa associated with the tumour microenvironment (Tahara *et al.*, 2014). Arguably the most consistent of these findings is that *Fusobacterium* species or more specifically *F. nucleatum* is enriched at the site of tumour tissue. A potential reason for this tumour tissue colonisation is via *F. nucleatum* Fap2 binding to Gal-GalNAc which is overexpressed in CRC. *Fusobacterium* isolates have also been cultured from liver metastases and share a high level of genome identity with *Fusobacterium* from the primary cancer

#### Chapter 1.7.5 *Bacteroides fragilis* toxin (BFT)

suggesting the bacterium can migrate with cancer cells to the metastatic site (Bullman *et al.*, 2017).

As with many microbes linked to cancer *F. nucleatum* can activate  $\beta$ -catenin signalling (Section 1.6.5). Two mechanisms behind this activation have been identified. The aforementioned FadA adhesin can bind to E-cadherin and induce nuclear translocation of  $\beta$ -catenin (Rubinstein *et al.*, 2013) while *F. nucleatum* lipopolysaccharide (LPS) can trigger TLR4 mediated  $\beta$ -catenin signalling (Chen *et al.*, 2017). In Apc<sup>Min/+</sup> mice *F. nucleatum* accelerated tumour onset and significantly increased the number of tumours compared with controls but did not stimulate colitis. *F. nucleatum* did not induce tumourigenesis or colitis in IL10<sup>-/-</sup> mice or T-bet<sup>-/-</sup> Rag2<sup>-/-</sup> mice (Kostic *et al.*, 2013).

Other oncogenic mechanisms linked to *F. nucleatum* include immune cell evasion and chemoresistance. The ability of its Fap2 protein to inhibit host immune cell killing of tumour cells (Gur *et al.*, 2015) and the activation of autophagy via Toll-like receptor 4 (TLR4) resulting in reduced apoptosis and therefore chemoresistance (Yu *et al.*, 2017).

#### **1.7.5 *Bacteroides fragilis* toxin (BFT)**

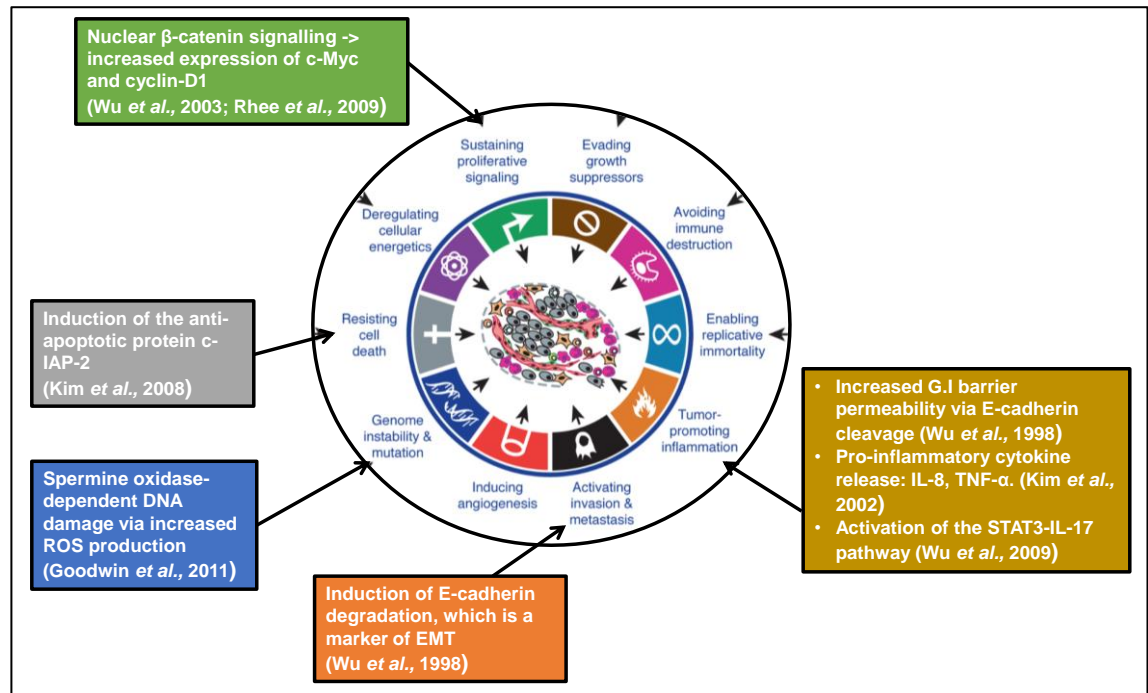
A subset of *B. fragilis* strains, termed Enterotoxigenic *B. fragilis* (ETBF), secrete a zinc metalloprotease known as BFT or fragilysin. A pivotal finding into the virulence potential of this toxin was made by injection of the purified protein into ligated animal intestinal loops. The result was a dose-dependent secretory response, tissue damage, crypt hyperplasia and neutrophil infiltration (Obiso *et al.*, 1995). The *bft* gene is contained on a 6kb pathogenicity island (BfPAI) (Franco, 2004) alongside another zinc metalloprotease (MP11) with 56% sequence similarity (Shiryaev *et al.*, 2014) within a conjugative transposon (Buckwold *et al.*, 2007). BFT can exist as one of three isotypes (bft-1, bft-2, bft-3) with whole genome sequencing revealing that non-toxigenic strains have likely acquired the BfPAI multiple times during evolution (Pierce and Bernstein, 2016). The majority of *B. fragilis* strains can be classed as either pattern I (contains BfPAI and transposon), pattern II (no BfPAI or transposon), or pattern III (transposon only) (Buckwold *et al.*, 2007).

#### Chapter 1.7.5 Bacteroides fragilis toxin (BFT)

BFT is a 397-residue protein with a signal peptide (18 residues), pro-domain (193 residue) and catalytic domain (181 residues). The catalytic domains of both BFT and MPII contain a zinc-binding consensus motif (HEXXHXXGXXH) and Met-turn characteristic of metzincin metalloproteinases. The 3D structure of both proteins have been determined and shows structural homology to human matrix metalloproteinases (MMPs) and a disintegrin and metalloproteinases (ADAMs) (Goulas *et al.*, 2011, Shiryayev *et al.*, 2014). The prodomain of both proteins is structurally similar and has a unique protein fold.

##### 1.7.5.1 BFT mechanism of action

The *bft* gene codes for a 44kDa preproprotein which when processed produces a 20kDa active toxin (Chung *et al.*, 1999). BFT is released into *B. fragilis* Outer Membrane Vesicles (OMVs) due to the formation of interactions with membrane components (Zakharzhevskaya *et al.*, 2017). It has been suggested that BFT binds to a currently unidentified membrane receptor (Wu *et al.*, 2006) or the OMVs are taken up by host eukaryotic cells (Zakharzhevskaya *et al.*, 2017) and subsequently trigger extensive signal transduction. BFT has been shown to induce cleavage and loss of the cell adhesion molecule, E-cadherin in HT-29 cells *in vitro* (Wu *et al.*, 1998) and in animal models (Rhee *et al.*, 2009) with rearrangement of the F-actin cytoskeleton (Saidi *et al.*, 1997). The adherens junction, located at cell-cell junctions, is a protein complex comprising of E-cadherin bound to and retaining  $\beta$ -catenin in the cytoplasm. Ectodomain shedding of E-cadherin by BFT is followed by  $\gamma$ -secretase dependent processing of the intracellular fragment (Wu *et al.*, 2007). The consequence of BFT-induced E-cadherin cleavage is IL-8 secretion (Hwang *et al.*, 2013) and nuclear localization of  $\beta$ -catenin (Figure 1-8). Once in the nucleus  $\beta$ -catenin can bind to transcription factors and induce pro-carcinogenic gene expression including c-Myc and cyclin D1 (Wu *et al.*, 2003).



**Figure 1-8 Pathogenic mechanisms of *B. fragilis* toxin (BFT)**

Oncogenic mechanisms of *B. fragilis* toxin and how they relate to the hallmarks of cancer. Figure adapted from (Hanahan and Weinberg, 2011).

#### 1.7.5.2 ETBF animal models

Animal models have been used to gain a greater understanding of the pathology of both acute and chronic ETBF infection. *Apc*<sup>Min/+</sup> mice ordinarily develop small intestine adenomas but when orally challenged with ETBF an earlier onset of mainly distally located tumours and inflammatory colitis is triggered (Housseau and Sears, 2010, Lu *et al.*, 2005b). The transcription factor signal transducer and activator of transcription-3 (STAT3) is rapidly activated which subsequently results in differentiation of T<sub>H</sub>17 cells and IL-17 secretion (Wu *et al.*, 2009, Wick *et al.*, 2014).

Wild type C57BL/6 mice infected with ETBF develop mucosal thickening, crypt abscesses, crypt hyperplasia and extensive neutrophilic infiltration whilst chronic ETBF infection results in persistent subclinical inflammation, enhanced epithelial hyperplasia and DNA damage (measured by activation of  $\gamma$ -H2A.x) via upregulation of spermine oxidase (SMO) and production of reactive oxygen species (Rhee *et al.*, 2009, Goodwin *et al.*, 2011, Brouwer *et al.*, 2013). ETBF colonization in either germ free mice or Mongolian gerbils gives rise to severe and consequently lethal colitis (Rhee *et al.*, 2009, Yim *et al.*, 2013)

## 1.8 Aims

The primary aim of this project is to investigate the contribution of the gastrointestinal commensal, *Bacteroides fragilis*, to gastrointestinal health and disease, with a specific focus on the multi-step process of colonic tumorigenesis.

### 1.8.1 Why *B. fragilis*?

The primary reason for studying *B. fragilis* is that some strains are potentially carcinogenic due to the ability to synthesise a tumourigenic toxin (BFT) (Wu *et al.*, 2003, Wu *et al.*, 2009, Housseau and Sears, 2010). *B. fragilis* as a species is a pathobiont and is ubiquitous in the G.I. tract. *B. fragilis* has recently been proposed as a probiotic species (Wang *et al.*, 2017) due to beneficial properties such as its ability to synthesise immunomodulatory capsular polysaccharides (Mazmanian *et al.*, 2005). The most studied of these polysaccharides is polysaccharide A (PSA), capable of inducing immune cell activation and preventing experimental colitis in mice (Mazmanian *et al.*, 2005). Bacterial capsules can also be considered a virulence factor as it is an immune system evasion strategy employed by the bacterium. *B. fragilis* is also a suitable organism for the study of host-microbe interactions as previous research suggests a predominately mucosal-associated organism. This property is shown by its ability to bind intestinal mucin *in vitro* (Huang *et al.*, 2011) and localisation within colonic crypts (Lee *et al.*, 2013).

The contribution of *B. fragilis* to gastrointestinal health is expected to involve both within-strain antigenic variation in response to environment cues, and differences in virulence potential between strains.

### 1.8.2 Objectives

- i. Sample collection
  - Establish a comprehensive sample collection comprising intestinal biopses, caecal fluid and blood samples from different patient cohorts with a focus on colon tumourigenesis
  - Establish robust methods for detection of relevant species and genes of interest

## Chapter 1.8.2 Objectives

### ii. Between-strain differences

- Determine the prevalence of *B. fragilis* as a species within samples collected from different patient cohorts and identify whether putative virulence genes are present in selected *B. fragilis* strains from these patient groups
- Given that multiple sequencing studies have identified *Fusobacterium* species as being enriched in the colorectal cancer microbiome, determination of the prevalence of *Fusobacterium* within these cohorts was undertaken to allow comparison with colorectal cancer microbiome sequencing studies

### iii. Within-strain differences

- Determine PSA promoter orientation (% on) within the different patient cohorts



## **Chapter 2 : MATERIALS AND METHODS**

### **2.1 Sample Collection**

#### **2.1.1 Subject Recruitment and Ethics Statement**

Participants undergoing routine diagnostic colonoscopy were recruited and samples collected at the Endoscopy Units at Guy's and St. Thomas NHS Hospital Foundation Trust. Ethical approval was sought from WALES REC 7 committee (REC number: 14/WA/1221) and Guy's & St Thomas' NHS Foundation Trust Research and Development (number: RJ115/N211). Informed written consent was obtained from all participants. Patient information sheet part 1 and 2 and the consent form are found in appendix 1 & 2. Exclusion criteria for recruitment included: under the age of 18, macroscopically active inflammatory bowel disease, antibiotics within the last 4 weeks, positive stool culture for the known enteropathogen *Clostridium difficile*. Individuals were divided into cohorts depending on the reason for attendance for, or result of colonoscopy. Subjects were recruited into newly diagnosed polyps/cancer (n = 32), previous polyps/cancer (n = 14), inflammatory bowel disease (n = 14) and healthy controls (n = 22) and are listed in tables 2.1 – 2.5. Some individuals were assigned multiple cohorts and consequently are listed separately in table 2.5. Healthy controls were defined as individuals with anaemia or rectal bleeding as an indication for colonoscopy with no signs of GI disease upon colonoscopy.

#### **2.1.2 Sample Collection and Processing**

##### **2.1.2.1 Mucosal Biopsies**

During colonoscopy mucosal biopsies were collected using 5 mm endoscopic biopsy forceps following completion to the caecum. Two mucosal biopsies from the ascending colon and two from the descending colon were collected and immediately frozen on dry ice. Similarly, two mucosal biopsies from the ascending colon and two from the descending colon were collected and immediately placed into RNeasy® (Sigma-Aldrich, Gillingham, UK) at room temperature. Biopsy samples collected from individuals with newly diagnosed tumours larger than 1 cm were collected on the tumour (x4 biopsies), 2 cm distal (x4 biopsies) and 10 cm distal (x4 biopsies). Samples stored in RNeasy® were kept at room temperature overnight then stored at -80 °C until used for analysis.

### **2.1.2.2 Caecal effluent**

Luminal fluid (20-30 mL) was collected from the caecum into a 40 mL Bard® mucus specimen trap (Bard Medical, Georgia, USA). Immediately following collection, the fluid was aliquoted into 3-4 x 15 mL Corning tubes with 50% filter sterilised glycerol. Samples were then frozen on dry ice and stored long term at -80 °C.

### **2.1.2.3 Blood**

A maximum of 26 mL blood was collected from each patient into relevant blood collection tubes and stored on ice packs in a cool bag until laboratory processing. All blood aliquots were stored long term at -80 °C.

- i. Whole blood: Approximately 8 mL blood was collected into 2 x 4 mL vacutainer EDTA blood collection tubes (BD™, Berkshire, UK). In the laboratory 1 mL aliquots of whole blood were made on ice inside a class II microbiological safety cabinet, type A2 (Nuaire).
- ii. Plasma: Approximately 8 mL blood was collected into 2 x 4 mL vacutainer EDTA blood collection tubes (BD™). Blood was centrifuged at 2000g for 10 min in a Microlite refrigerated centrifuge (Thermo-Fisher Scientific, Nottingham, UK) at 4°C. Aliquots (500 µL) were made on ice inside a class II microbiological safety cabinet (Nuaire, Caerphilly, Wales).

Serum: Approximately 10 mL blood was collected into a vacutainer SS II Advance Tube with gold hemogard closure (BD™). The clot was removed by centrifugation at 2000g for 10 min in a Microlite refrigerated centrifuge (Thermo-Fisher Scientific) at 4°C. Aliquots (500 µL) were made on ice inside a class II microbiological safety cabinet (Nuaire).

The relevant clinical data was also obtained (see Tables 2.1 – 2.5 on pages 60-69).

**Table 2-1 Healthy patient cohort**

Patient ID	Biopsies	Caecal Effluent	Blood	Age at collection	Sex	Indication for colonoscopy	Other medical history
BF012	AC, DC	N/A	N/A	77	M	Anaemia	N/A
BF018	AC, DC	3 x aliquots	N/A	49	M	Rectal bleeding	N/A
BF022	AC, DC	3 x aliquots	8 x Whole blood, 8 x Plasma, 8 x Serum aliquots	41	M	Anaemia	Type 1 diabetes, OGD also clear
BF031	AC, DC	4 x aliquots	8 x Whole blood, 8 x Plasma, 8 x Serum aliquots	21	F	Rectal bleeding and anaemia	Cryoglobulinemic vasculitis, membranoproliferative glomerulonephritis
BF036	AC, DC	4 x aliquots	8 x Whole blood, 8 x Plasma, 8 x Serum aliquots	79	F	Anaemia	Ischemic heart disease, triple bypass, hypertension, type II diabetes
BF046	AC, DC	4 x aliquots	8 x Whole blood, 8 x Plasma, 12 x Serum aliquots	57	M	Rectal bleeding, altered bowel habit, loss of weight	Bladder cancer transurethral resection
BF051	AC, DC	4 x aliquots	8 x Whole blood, 9 x Plasma, 8 x Serum aliquots	65	F	Rectal bleeding, family history of CRC	Colonoscopy found small rectal leiomyoma On statins and thyroxine
BF052	AC, DC	3 x aliquots	8 x Whole blood, 8 x Plasma, 7 x Serum aliquots	53	M	Rectal bleeding	N/A
BF057	AC, DC	4 x aliquots	N/A	51	F	Rectal bleeding, altered bowel habit	Found an ulcer in the ileocaecal valve and DC. No inflammation at the site of the biopsies
BF063	AC, DC	3 x aliquots	8 x Whole blood, 7 x Plasma, 7 x Serum aliquots	79	M	Rectal bleeding	Pacemaker

BF064	AC, DC	4 x aliquots	9 x Whole blood, 8 x Plasma, 8 x Serum aliquots	57	F	Anaemia	N/A
BF065	AC, DC	1 x aliquot	8 x Whole blood, 7 x Plasma, 9 x Serum aliquots	49	F	Rectal bleeding, family history of CRC	N/A
BF074	AC, DC	4 x aliquots	7 x Whole blood, 3 x Plasma, 5 x Serum aliquots	55	M	Anaemia	Diverticular, OGD <i>H. pylori</i> positive
BF077	AC, DC	N/A	11 x Whole blood, 8 x Plasma, 6 x Serum aliquots	25	M	Rectal bleeding	N/A
BF086	AC, DC	3 x aliquots	8 x Whole blood, 7 x Plasma, 7 x Serum aliquots	49	M	Anaemia	Diverticular disease
BF088	AC, DC	3 x aliquots	7x Whole blood, 7 x Plasma, 5 x Serum aliquots	47	F	Rectal bleeding	Diverticular disease and haemorrhoids
BF095	AC, DC	4 x aliquots	N/A	62	M	Rectal bleeding	N/A
BF106	AC, DC	3 x aliquots	6 x Whole blood, 6 x Plasma, 6 x Serum aliquots	22	F	Rectal bleeding, diarrhoea, mucous	N/A
BF111	AC, DC	3 x aliquots	6 x Whole blood, 5 x Plasma, 7 x Serum aliquots	61	F	Anaemia	Diverticular disease
BF121	AC, DC	3 x aliquots	6 x Whole blood, 6 x Plasma, 5 x Serum aliquots	59	M	Rectal bleeding	CVA, type 2 diabetes, hypercholesterol
BF123	AC, DC	3 x aliquots	7 x Whole blood, 5 x Plasma, 5 x Serum aliquots	35	F	Rectal bleeding	N/A

## Chapter 2.1.2 Sample Collection and Processing

BF124	AC, DC	3 x aliquots	N/A	49	F	Rectal bleeding	N/A
-------	--------	--------------	-----	----	---	-----------------	-----

AC = Ascending colon; DC = Descending colon; OGD = oesophago-gastroduodenoscopy; CVA = Cerebrovascular accident

**Table 2-2 Inflammatory Bowel Disease Patient Cohort**

Patient ID	Biopsies	Caecal Effluent	Blood	Age at collection	Sex	Histopathology Results	Other medical history
BF001	AC, DC	N/A	8 x Whole blood, 7 x Plasma, 8 x Serum aliquots	54	F	Biopsies showed no inflammation	Ulcerative colitis (UC)
BF003	AC, DC	N/A	N/A	42	F	Mild patchy inflammation with focal active inflammation	Colonic Crohn's, angina, hypertension
BF015	AC, DC	3 x aliquots	8 x Whole blood, 6 x Plasma, 7 x Serum aliquots	51	M	Biopsies showed mild-moderate active, chronic colitis	Colonic Crohns, pancolitis and primary sclerosing cholangitis (PSC)
BF020	AC,DC	3 x aliquots	8 x Whole blood, 7 x Plasma, 6 x Serum aliquots	36	F	Currently inactive colitis	Previous surgery for colitis
BF033	AC,DC	4 x aliquots	8 x Whole blood, 7 x Plasma, 7 x Serum aliquots	34	M	No inflammation	Unclassified UC
BF070	AC,DC	3 x aliquots	8 x Whole blood, 6 x Plasma, 4 x Serum aliquots	37	M	Inactive	Severe flare March 2015, multiple post-inflammatory pseudopolyps
BF084	AC,DC	4 x aliquots	8 x Whole blood, 8 x Plasma, 7 x Serum aliquots	54	M	AC biopsy taken near anastomosis, biopsies showed some inflammation	Pan-enteric Crohn's, previous right hemicolectomy
BF099	AC,DC	3 x aliquots	7 x Whole blood, 6 x Plasma, 6 x Serum aliquots	75	F	N/A	Diverticular disease, Colonic Crohn's
BF117	AC,DC	3 x aliquots	7 x Whole blood, 6 x Plasma, 4 x Serum aliquots	41	M	N/A	Pan - UC

AC = Ascending colon; DC = Descending colon

**Table 2-3 Previous polyps/cancer patient cohort**

Patient ID	Biopsies	Caecal Effluent	Blood	Age at collection	Sex	Indication for colonoscopy	Other medical history
BF016	AC,DC	3 x aliquots	7 x Whole blood, 8 x Plasma, 8 x Serum aliquots	64	M	Polyp surveillance	Previous right sided hemicolectomy
BF017	AC,DC	N/A	N/A	59	M	Polyp surveillance	Prostrate cancer
BF024	AC,DC	4 x aliquots	8 x Whole blood, 8 x Plasma, 7 x Serum aliquots	72	F	Changes in bowel habits and weight loss	N/A
BF042	AC,DC	4 x aliquots	8 x Whole blood, 8 x Plasma, 5 x Serum aliquots	76	F	Polyp surveillance, change in bowel habits	Previous rectal polyp 3 years ago (tubular adenoma), rheumatoid arthritis, diverticular disease
BF045	AC,DC	4 x aliquots	8 x Whole blood, 8 x Plasma, 5 x Serum aliquots	57	M	CRC surveillance	Previous adenocarcinoma, anterior resection
BF059	AC,DC	4 x aliquots	8 x Whole blood, 9 x Plasma, 3 x Serum aliquots	51	F	Polyp surveillance	Previous pedunculated distal, sigmoid polyp
BF060	AC,DC	N/A	8 x Whole blood, 8 x Plasma, 6 x Serum aliquots	73	F	CRC surveillance	Previous adenocarcinoma, right sided hemicolectomy
BF071	AC,DC	4 x aliquots	N/A	59	M	Recent rectal bleeding	Previous anal cancer, HIV positive
BF076	AC,DC	4 x aliquots	7 x Whole blood, 8 x Plasma, 7 x Serum aliquots	65	M	Polyp surveillance	Previous polyp (2011 tubulovillous adenoma)
BF080	AC,DC	N/A	N/A	75	M	Family history CRC, change in bowel habit	Previous polyp, diverticular disease, COPD, gout, hypertension
BF091	AC,DC	3 x aliquots	6 x Whole blood, 6 x	72	M	CRC surveillance	Previous moderately differentiated

## Chapter 2.1.2 Sample Collection and Processing

			Plasma, 5 x Serum aliquots				adenocarcinoma pT3a N0, anterior resection
BF097	AC,DC	3 x aliquots	7 x Whole blood, 6 x Plasma, 6 x Serum aliquots	43	M	Family history CRC	Previous polyp with low grade dysplasia (transverse colon, tubular adenoma)
BF098	AC,DC	4 x aliquots	8 x Whole blood, 8 x Plasma, 7 x Serum aliquots	81	F	CRC surveillance	Previous CRC, transverse colon 2014

AC = Ascending colon; DC = Descending colon; COPD = Chronic obstructive pulmonary disease



**Table 2-4 Newly diagnosed polyp/cancer patient cohort**

Patient ID	Biopsies	Caecal Effluent	Blood	Age at collection	Sex	Colonoscopy/Histopathology Results	Other medical history
BF002	OT, NT, AT	N/A	N/A	73	M	Tubular adenoma	Right hemicolectomy due to previous adenocarcinoma, type 2 diabetes, haemorrhoids
BF005	OT, NT, AT	2 x aliquots	8 x Whole blood, 7 x Plasma, 8 x Serum aliquots	54	M	N/A	N/A
BF006	OT, NT, AT	N/A	N/A	91	M	N/A	N/A
BF007	OT, NT, AT	N/A	N/A	72	F	N/A	N/A
BF008	OT, NT, AT	N/A	N/A	71	F	Transverse serrated sessile polyp and 2cm sessile lesion (biopsies taken from) is adenocarcinoma T3(b) No nodes Dukes B, proximal AC	Previous polyp, breast cancer, previous skin cancer
BF009	AC, DC	N/A	N/A	73	F	Found 3 new polyps in caecum, 2x sigmoid and 1 descending colon	N/A
BF011	AC, DC	N/A	N/A	69	M	Multiple new small polyps, 1 large in rectum and flat hyperplastic polyp in caecum	Previous adenomas and serrated lesions, raised cholesterol
BF013	OT, AC	N/A	N/A	N/A	M	Unknown	Previous polyps in AC and caecum, hypertension
BF023	AC, DC	4 x aliquots	8 x Whole blood, 7 x Plasma, 8 x Serum aliquots	68	M	Small caecal polyp	Iron deficient anaemia, Metastatic, prostate cancer
BF027	OT, NT, AT	4 x aliquots	8 x Whole blood, 4 x Plasma, 2 x Serum aliquots	36	M	Pedunculated polyp in DC- inflammatory pseudopolyp	Change in bowel habit, weight loss, Diabetic renal transplant
BF028	OT, NT, AT	3 x aliquots	N/A	70	F	Sessile, flat hyperplastic polyps. Biopsies taken from sigmoid colon. Previous	Squamous cell carcinoma on leg,

						tubuloadenoma and hyperplastic polyps	
BF029	OT, NT, AT	4 x aliquots	8 x Whole blood, 8 x Plasma, 5 x Serum aliquots	76	F	3 small sessile polyps in descending colon, biopsies taken from large pedunculated tubulovillous adenoma with low grade dysplasia in sigmoid colon.	N/A
BF030	AC, DC	4 x aliquots	8 x Whole blood, 6 x Plasma, 6 x Serum aliquots	51	M	Presented with rectal bleeding. Hyperplastic polyp in the distal transverse colon. 2 tubular adenomas with low grade dysplasia in sigmoid colon	N/A
BF032	AC, DC	4 x aliquots	1 x Whole blood, 7 x Plasma, 6 x Serum aliquots	74	F	Caecal polyp	N/A
BF034	OT, NT, AT	N/A	8 x Whole blood, 8 x Plasma, 8 x Serum aliquots	54	F	Pedunculated polyp - tubular adenoma low grade dysplasia	HIV positive, varices, portal vein thrombosis
BF043	AC, DC	3 x aliquots	8 x Whole blood, 8 x Plasma, 8 x Serum aliquots	74	M	Presented with persistent anaemia and change in bowel habit. Inflammatory polyp in sigmoid colon	N/A
BF044	AC	4 x aliquots	8 x Whole blood, 9 x Plasma, 8 x Serum aliquots	61	M	Presented with rectal bleeding and previous polyps. Tubulovillous adenoma in AC. Unknown previous polyps histology	Type 2 diabetes, failed renal transplant
BF047	AC, DC	3 x aliquots	8 x Whole blood, 9 x Plasma, 8 x Serum aliquots	37	M	Small hyperplastic rectal polyp. Previous rectal polyps, Family history of CRC	N/A
BF048	OT, NT, AT	N/A	8 x Whole blood, 8 x Plasma, 7 x Serum aliquots	45	M	Large sigmoid polyp, 2 small polyps distal to larger polyp	N/A
BF054	AC, DC	N/A	8 x Whole blood, 8 x Plasma, 8 x Serum aliquots	59	M	Small tubular adenoma descending colon; previous sigmoid polyps tubulovillous low-grade dysplasia	N/A
BF058	AC, OT	4 x aliquots	8 x Whole blood, 8	42	M	Previous polyp removed from ascending	N/A

# Chapter 2.1.2 Sample Collection and Processing

			x Plasma, 4 x Serum aliquots			colon. New polyp.	
BF061	OT, NT, AT	4 x aliquots	8 x Whole blood, 8 x Plasma, 6 x Serum aliquots	76	M	Presented with rectal bleeding and constipation. Rectosigmoid cancer	Thyroid disease and osteoarthritis
BF068	OT, DC	N/A	8 x Whole blood, 5 x Plasma, 6 x Serum aliquots	82	M	Presented with diarrhoea and anaemia. Severe splenic stricture, probable cancer	Previous hemicolectomy due to adenomatous polyp
BF071	AC, DC	3 x aliquots	N/A	59	M	Presented with rectal bleeding	Previous anal cancer, HIV positive
BF072	OT, NT, AT	N/A	8 x Whole blood, 6 x Plasma, 5 x Serum aliquots	48	M	Presented with diarrhoea and weight loss. Small sigmoid polyp	Diabetes
BF079	NT, AT	N/A	10 x Whole blood, 8 x Plasma, 6 x Serum aliquots	76	M	Presented with anaemia and previous polyps. New polyps which were sessile serrated, hyperplastic and tubular adenomas	Diabetes, renal disease, COPD, hypertension
BF090	AC, DC	4 x aliquots	8 x Whole blood, 7 x Plasma, 4 x Serum aliquots	73	F	1cm rectal tubular adenoma	Diverticular disease, hypertension
BF100	AC, DC	4 x aliquots	2 x Whole blood,	59	F	Presented with rectal bleeding. Small polyp in ascending colon. Tubular adenoma in rectum	Recent lung cancer and previous breast cancer
BF115	OT, NT, AT	3 x aliquots	N/A	73	M	Tubulovillous adenoma sigmoid colon.	Deceased March 2017. Metastatic right renal cell carcinoma, ischaemic heart disease

AC = Ascending colon; DC = Descending colon; OT = On tumour; NT = Next to tumour; AT = Away from tumour

**Table 2-5 Patients assigned multiple cohorts**

Patient ID	Biopsies	Caecal Effluent	Blood	Age at collection	Sex	Colonoscopy/Histopathology Results and other medical history	Patient groups assigned
BF014	AC, DC	N/A	N/A	77	M	No new polyps or inflammation	IBD and previous polyps/cancer
BF019	OT, NT, AT	N/A	8 x Whole blood, 6 x Plasma, 7 x Serum aliquots	61	M	2 x tubulovillous polyps with low grade dysplasia. 1 in hepatic flexure, 1 in sigmoid colon (~2 cm). Biopsies taken from sigmoid polyp.	IBD and newly diagnosed polyp/cancer
BF035	AC, DC	4 x aliquots	8 x Whole blood, 8 x Plasma, 8 x Serum aliquots	63	M	Long-term IBD (UC), pseudopolyps, abnormal areas in ileocaecal valve, transverse and sigmoid colon. High grade dysplasia adenocarcinoma found in sigmoid colon. Primary sclerosing cholangitis, varices	IBD and newly diagnosed polyp/cancer
BF041	AC, DC	4 x aliquots	8 x Whole blood, 7 x Plasma, 7 x Serum aliquots	25	M	Inflammatory polyps in caecum. Crohn's in remission	IBD and newly diagnosed polyp/cancer
BF049	OT, NT, AT	N/A	N/A	49	M	Tubulovillous adenoma in rectum. Ileocolonic Crohn's –no inflammation	IBD and newly diagnosed polyp/cancer
BF056	AC, DC	4 x aliquots	N/A	70	F	Distal UC. Small rectal polyp, no inflammation. Previous tubulovillous adenoma with low grade dysplasia in sigmoid colon.	IBD and newly diagnosed polyp/cancer

AC = Ascending colon; DC = Descending colon; OT = On tumour; NT = Next to tumour; AT = Away from tumour

## **2.2 DNA Processing**

### **2.2.1 DNA Extraction from Intestinal Biopsies**

DNA was extracted from intestinal biopsy samples using the FastDNA SPIN kit for soil (MP Biologicals, California, USA) following the manufacturer's instructions. DNA extraction included a bead-beating step using a FastPrep24 (MP Biologicals) homogenizer. Other DNA extraction steps were carried out in a Class II microbiological safety cabinet (Nuair). DNA aliquots were placed into 0.5 mL DNase free DNA LoBind® microcentrifuge tubes (Sigma-Aldrich). Extracted DNA aliquots were stored at -20°C until further processing, while extracted DNA was refrigerated short-term at 4°C. Extracted DNA was initially subjected to PCR using universal primers to determine suitability of the DNA for further PCR analysis.

### **2.2.2 DNA Quantification**

DNA was quantified by a fluorescence-based method using the Qubit™ double-stranded DNA high sensitivity assay kit (Life Technologies, California, USA) and Qubit™ 3.0 Fluorometer 3.0 (Life Technologies). Quantification was carried out according to the manufacturer's instructions.

## **2.3 End-point Polymerase Chain Reaction (PCR)**

### **2.3.1 Primers**

PCR primers were ordered from Eurofins Genomics (Ebersburg, Germany) and obtained in a lyophilised form. Primers were rehydrated in sterile molecular grade water (Sigma-Aldrich) to a concentration of 100 pmol/μL. Primer aliquots at a concentration of 10 pmol/μL were made up ready for use in PCR and stored long term at -20°C and short term at 4°C.

### **2.3.2 Polymerase Chain Reaction (PCR)**

PCR reactions were performed in a final volume of 25 μL with each reaction mixture containing: 12.5 μL 2X DreamTaq™ polymerase mastermix (Fermentas, Waltham, USA), 10.5 μL sterile molecular grade H<sub>2</sub>O (Sigma-Aldrich), 0.5 μL forward primer 10 pmol/μL (Eurofins genomics), 0.5 μL reverse primer (10 pmol/μL) (Eurofins genomics) and 1 μL of template DNA. All PCR reactions were normalised to use 40 ng of extracted biopsy DNA. PCR reactions were carried out in a TC 412 thermal cycler using cycling conditions of an initial denaturation step of 5 min at

### Chapter 2.3.3 Colony PCR

95°C; 30 cycles of amplification consisting of denaturation (45 s at 95°C), annealing (45 s at X°C) and extension (X at 72°C); and a final extension of 15 min at 72°C where X is variable for each PCR target. A no DNA template, negative control was included in every PCR.

#### **2.3.3 Colony PCR**

Colony PCR in this thesis refers to PCR performed on bacterial DNA without a specific DNA extraction step. A single bacterial colony was collected into 50 µL sterile molecular grade H<sub>2</sub>O (Sigma-Aldrich) and 1 µL of this used as a DNA template per PCR reaction. An initial denaturation step of the PCR programme (5 min at 95°C) was used to lyse bacterial cells and release the DNA. PCR reactions were carried out using thermal cycling conditions as described in section 2.3.2.

#### **2.3.4 DNA Agarose Gel Electrophoresis**

PCR amplicons and extracted DNA were visualized on a 1-2% (w/v) agarose gels. Agarose gels were made using molecular grade agarose (Sigma-Aldrich) dissolved in 0.5X Tris-Borate EDTA (TBE) buffer (VWR) containing 0.01% (v/v) GelRed™ nucleic acid gel stain (VWR). Individual samples were mixing with 6X loading dye (Thermo-Scientific, Nottingham, UK) and visualised on the Alphamager HP Gel Imaging System (Alpha Innotech, California, USA). The molecular size of the DNA fragments were estimated using either GeneRuler Express DNA ladder (Thermo-Fisher Scientific), GeneRuler 100 bp Plus DNA ladder (Thermo-Fisher Scientific) or Lambda DNA-Hind III Digest (New England Biolabs) molecular markers.

#### **2.3.5 PCR Amplicon Purification and Sequencing**

The method of PCR amplicon purification depended on whether one or multiple DNA fragments were present. If multiple bands were present the fragment of interest was excised from the gel and purified using the QIAquick PCR purification kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. If a single band was present the PCR product was purified using the CleanSweep™ PCR purification kit (Thermo-Fisher Scientific). CleanSweep™ reagent (2 µL) was added to 5 µL PCR product and incubated at 37°C for 15 min followed by 15 min incubation at 80°C using the PCR thermocycler. Purified PCR amplicons were sent for DNA sequencing using Eurofins Genomics TubeSeq service.

## **2.4 Protein techniques**

### ***2.4.1 Protein Extraction from Intestinal Biopsy Samples***

Single intestinal biopsies stored in RNeasy® were used for simultaneous RNA (data not shown) and protein extraction. Excess RNeasy® was removed and the biopsy divided in half. Ice-cold RIPA lysis and extraction buffer (250 µL) (Thermo Fisher) was added to the half of the biopsy allocated for protein extraction. The sample was then incubated on ice for 20-30 min. The sample was centrifuged at 10,000 g for 15 min at 4°C before being aliquoted and stored at -80°C until analysis.

### ***2.4.2 Total protein quantification***

Protein was quantified by a fluorescence-based method in duplicate using the Qubit™ protein assay kit (Life Technologies) and Qubit™ 3.0 Fluorometer 3.0 (Life Technologies) according to the manufacturer's instructions.

### ***2.4.3 Enzyme-Linked Immunosorbant Assay (ELISA)***

The human IL-10 ELISA kit (Invitrogen) was used to quantify IL-10 concentrations in serum and intestinal biopsy tissue lysates. 15 µL of biopsy lysate diluted in 35 µL standard diluent buffer or 50 µL of serum were added in duplicate. Following a 2 hour incubation at room temperature the 96 well plate was washed four times using 1X wash buffer. 100 µL Hu IL-10 biotin conjugate was added to all wells except chromogen blanks and incubated for 2 hours at room temperature. The plate was washed four times with 1 x wash buffer and 100 µL streptavidin-HRP solution was added to all wells except chromogen blanks. The plate was incubated for 30 minutes and washed four times with 1 x wash buffer. 100 µL stabilized chromogen was added to the plate and incubated for 30 minutes at room temperature in the dark. The reaction was stopped using 100 µL stop solution and the plate read at an absorbance of 450 nm.

Tissue IL-10 concentration in pg/mL was normalized against total protein, as determined by Qubit protein assay, i.e. ELISA IL-10 concentration (pg/mL) / total protein concentration (mg/mL) to obtain a final result in terms of pg IL-10 per mg protein. Serum and tissue IL-10 concentrations were not normally distributed so non-parametric statistical tests were used. Comparison of medians amongst all patient groups was carried out using the Kruskal-Wallis

test. Dunn's multiple comparisons test was used to compare means between pairs of patient groups. Analysis of correlation between IL-10 and B. fragilis PSA orientation was tested using Spearman's rank correlation.

## 2.5 Microbiology

### 2.5.1 Microbiology

Bacterial reference strains were purchased either from Thermo Scientific Oxoid or the Japan Collection of Microorganisms (JCM), RIKEN BioResource Center, Ibaraki, Japan. Bacterial reference species and strains are detailed in Tables 1-2–1-4.

**Table 2-6 Table of Bacterial Reference Strains**

\* = Type strain

Species	Strain	Original Source of Isolate	Supplier
<i>Bacteroides ovatus</i> (Eggerth and Gagnon, 1933)	ATCC® 8483™*(Eggerth and Gagnon, 1933)	Human faeces	Culti-Loop™, Thermo-Scientific
<i>Bacteroides uniformis</i> (Eggerth and Gagnon, 1933)	ATCC® 8492™	Human faeces	
<i>Bacteroides thetaiotaomicron</i>	ATCC® 29741™	Perforated appendix	
<i>Bacteroides salyersae</i> (Song <i>et al.</i> , 2004)	ATCC® BAA-997™* (JCM 12988)	Human appendix tissue	Japan Collection of Microorganisms
<i>Bacteroides sartorii</i> (Clavel <i>et al.</i> , 2010)	JCM 17136*	Mouse caecal sample	
<i>Bacteroides coprophilus</i> (Hayashi <i>et al.</i> , 2007)	JCM 13818*	Human faeces	
<i>Prevotella melaninogenica</i>	ATCC® 25845™*	Sputum	Culti-Loop™, Thermo-Scientific
<i>Fusobacterium mortiferum</i>	ATCC® 9817™	Unknown	
<i>Fusobacterium periodonticum</i>	ATCC® 33693™	Periodontitis	Kind gift from Professor William Wade



**Table 2-7 Table of *Bacteroides fragilis* Reference Strains**

\* = Type strain

<i>Bacteroides fragilis</i> Strain	<i>B. fragilis</i> toxin (BFT)	Original Source of Isolate	Supplier
NCTC 9343/ ATCC 25285*	Non-toxigenic	Human appendix abscess	Culti-Loop™, Thermo-Scientific
JCM 17585 (Odamaki <i>et al.</i> , 2008)	BFT Subtype 1	Human faeces	Japan Collection of Microorganisms
JCM 17586 (Odamaki <i>et al.</i> , 2008)	BFT Subtype 2	Human faeces	
JCM 17587 (Odamaki <i>et al.</i> , 2008)	BFT Subtype 3	Human faeces	

**Table 2-8 Table of *Fusobacterium nucleatum* Reference Strains**

\* = Type strain

<i>Fusobacterium nucleatum</i> subspecies and strain	Strain features	Original Source of Isolate	Supplier
<i>nucleatum</i>	ATCC® 25586™*	Cervico-facial lesion	Culti-Loop™, Thermo-Scientific
<i>polymorphum</i>	ATCC® 10953™*	Inflamed gingiva	
<i>animalis</i> (Gharbia and Shah, 1992)	ATCC® 51191™ (JCM 11025*)	Unknown	Japan Collection of Microorganisms
<i>vincentii</i> (Dzink <i>et al.</i> , 1990)	ATCC® 49256™ (JCM 11023)	Human oral cavity	
<i>fusiforme</i> (Gharbia and Shah, 1992)	NCTC 11326	Human sinusitis upper jaw	Kind gift from Professor William Wade

### 2.5.2 Culturing Conditions

Reference Culti-Loop™ strains were streaked onto reduced Fastidious Anaerobe Agar (Lab M Limited, Bury, UK) with 5% (v/v) defibrinated horse blood (Thermo-Fisher Scientific). Strains obtained from JCM were supplied as lyophilized bacterial cultures and rehydrated using nutrient broth. Nutrient broth was prepared using 0.5% w/v Bacto Peptone (Becton Dickinson, New Jersey, USA), 0.3% w/v beef extract (Sigma-Aldrich), pH 7.0. Bacteria were cultured in a MACS-MG-1000 anaerobic workstation (Don Whitley, Shipley, UK) with an atmosphere of 80% N<sub>2</sub>, 10% H<sub>2</sub> and 10% CO<sub>2</sub> at 37°C.

### Chapter 2.5.3 Plate wash method to detect bacterial isolates

For long-term storage strains were cultured in either pre-reduced Brain-Heart Infusion (BHI) Broth (VWR International, Pennsylvania, USA) or Tryptone Soya Broth (VWR International) and frozen with 10% (v/v) glycerol (VWR) at -80°C.

#### **2.5.3 Plate wash method to detect bacterial isolates**

Caecal effluent (500 mL) sample was used to make 3-4 10-fold serial dilutions. Dilutions were made using 50 µL caecal effluent diluted in 450 µL pre-reduced tryptone soya broth. 100 µL each dilution were spread in duplicate onto *Bacteroides* Selective Agar and *Fusobacterium* Selective Agar. The *Bacteroides* Selective Agar used was Schaedler Agar (Sigma-Aldrich) supplemented with 100 mg/L kanamycin (Santa Cruz Biotechnology, Texas, USA), 7.5 mg/L vancomycin (Sigma-Aldrich) and 5% defibrinated sheep blood (Thermo-Fisher Scientific). The agar selective for *Fusobacterium* was Schaedler Agar supplemented with 5 mg/L crystal violet, 4 mg/L erythromycin and 5% defibrinated sheep blood. The agar plates were cultured for 72 h in an anaerobic environment. One plate of each dilution was washed using 200 µL sterile water and 'colony' PCR (section 2.3.3) using relevant primers carried out. Plates with positive results were used to pick 10-20 colonies, which were restreaked onto Fastidious Anaerobic Agar supplemented with 5% (v/v) defibrinated horse blood. 'Colony' PCR was used to detect the presence of any *B. fragilis* or *Fusobacterium* isolates and positive strains were cultured in TSB to make glycerol stocks as described previously.

## **2.6 Polysaccharide A (PSA) Promoter Orientation**

PSA promoter orientation was determined using a method adapted from (Krinos *et al.*, 2001; Troy *et al.*, 2010). PCR of a fragment of DNA including the PSA promoter was carried out using the PSAF1 and PSAF2 primers (Table 2-11). PCRs were set up as described in section 2.3.2 using an annealing temperature of 57 °C and an extension time of 70 s. A restriction digest of the PCR amplicons was set up in a 1.5 mL microcentrifuge tube using 19 µL sterile, DEPC treated water (Sigma-Aldrich), 8 µL PCR product DNA, 2 µL FastDigest green buffer (Thermo-Fisher Scientific) and 1 µL FastDigest SspI enzyme (Thermo-Fisher Scientific). Restriction digests were incubated for 20 min at 37°C. 8 µL restriction digest was loaded and visualised using DNA agarose gel electrophoresis as described in section 2.3.4. Agarose gels were visualised using the ChemiDoc™ MP UV transilluminator (Bio-Rad Laboratories) and DNA band

intensity determined using the Image Lab (version 5.2.1) software (Bio-Rad Laboratories). The expected sizes of the undigested and digested fragments are detailed in Table 2-10.

**Table 2-9 PCR digestion fragment sizes**

Product size (bp)	Fragment sizes (bp) ON		Fragment sizes (bp) OFF	
1076	257	819	148	928

The percentage of the *B. fragilis* population with the PSA promoter oriented on was calculated as follows:

$$= (\text{Total band intensity of 'ON' fragments} / \text{Total band intensity of all fragments}) \times 100$$

Non-parametric statistical comparison of medians amongst all patient groups was carried out using the Kruskal-Wallis test. Dunn's multiple comparisons test was used to compare means between pairs of patient groups.

**Table 2-10 Polysaccharide A Promoter Region Primers**

Primer name	Primer Sequence (5'-3')	GC content %	Tm (°C)	Size (bp)
PSAF1	TGTGTAAATGATAGGAGGCTAGGG	61.0	45.8	24
PSAR1	GTTGACGGAAATGATCGGTATAG	58.9	43.5	23

## 2.7 Statistical Analysis

Statistical analysis was performed using Prism v. 7.0 for Mac OS (GraphPad Software, San Diego, California). Determination of significant differences amongst patient groups in terms of prevalence of bacterial taxa and strains/genes of interest was determined using the Fisher's exact test.

## **Chapter 3 : END-POINT POLYMERASE CHAIN REACTION DEVELOPMENT**

### **3.1 Introduction and Aims**

Advances in next generation sequencing technology and a subsequent reduction in cost have resulted in a large expansion of microbiome research. This technology has provided a wealth of in-depth data and progressed knowledge of bacterial diversity within the G.I. tract, including previously undescribed and as yet unculturable species. In regards to the CRC microbiome many studies have been carried out and their key findings are summarised in detail in section 1.7. Unfortunately, the data sets from microbiome studies often contain insufficient information to look at strain level detail to test the hypothesis that BFT is associated with human CRC. There are therefore limitations to this large data approach and a more specific methodology was sought and optimised. Previous studies specifically investigating BFT epidemiology (Toprak *et al.*, 2006, Boleij *et al.*, 2015) have limitations, particularly in respect to a lack of a suitable healthy cohort or sample type. This chapter describes the development of a PCR assay to robustly investigate the prevalence of *Fusobacterium* and its associated virulence gene *fadA* (section 1.7.4), and the species *B. fragilis*, and its putative virulence genes *bft* (section 1.7.5) and *ubb* (section 1.5.2.1). A comprehensive collection of colonic tissue samples from various patient cohorts was collected to probe the association of the aforementioned bacterial taxa and genes with gastrointestinal disease, and in particular with colon tumorigenesis.

#### **3.1.1 Methods of bacterial identification and characterisation**

There is no gold standard method for bacterial detection and/ or identification, but all methods can be classed as culture-dependent or culture-independent. Results using both methods are described in this thesis but primarily a culture-independent method was used and its development is described within this chapter.

Culture-dependent methods grow viable bacteria from a sample with subsequent identification based on morphology, phenotypic characterisation or sequencing. Shortcomings of this technique are that some species have no known method of culture and bacteria have mechanisms by which they can outcompete each other or exchange genetic material.

### Chapter 3.1.2 Polymerase Chain Reaction

Culture-independent methods include the sequencing of DNA or cDNA which allows the identification of both non-viable and viable bacteria respectively. This technique can be affected by contamination, results may vary depending on the method of DNA extraction used (Kennedy *et al.*, 2014) and DNA sequencing does not discriminate between viable and non-viable organisms.

#### 3.1.1.1 Bacterial identification using the 16S rRNA gene

A common sequencing method of bacterial identification uses taxonomy dependent differences in the 16S rRNA gene. This gene codes for 16S ribosomal RNA in bacteria and archaea and is a component of the 30S small subunit of prokaryotic ribosomes. An early finding of DNA sequencing of 16S rRNA and eukaryotic 18S rRNA genes led to the realisation that rRNA sequences of various organisms could be used to determine phylogenetic relationships. Indeed Carl Woese used this insight to identify three domains of life: Archaea, Bacteria, and Eukarya whereas previously only prokaryota and eukaryota were described (Woese *et al.*, 1990).

This housekeeping gene is commonly used for bacterial identification in part because it is an ancient gene with an unchanged function and therefore sequence changes are likely to be the result of gradual evolution over time. The primary reason for the use of this gene for bacterial identification is that regions of the sequence are universally conserved across species while other regions are taxa-specific including some species-specific regions. Another motivation for the use of this housekeeping gene is that comprehensive quality-controlled databases of full-length 16S rRNA sequences exist (Maidak *et al.*, 1999, Quast *et al.*, 2013).

#### 3.1.2 Polymerase Chain Reaction

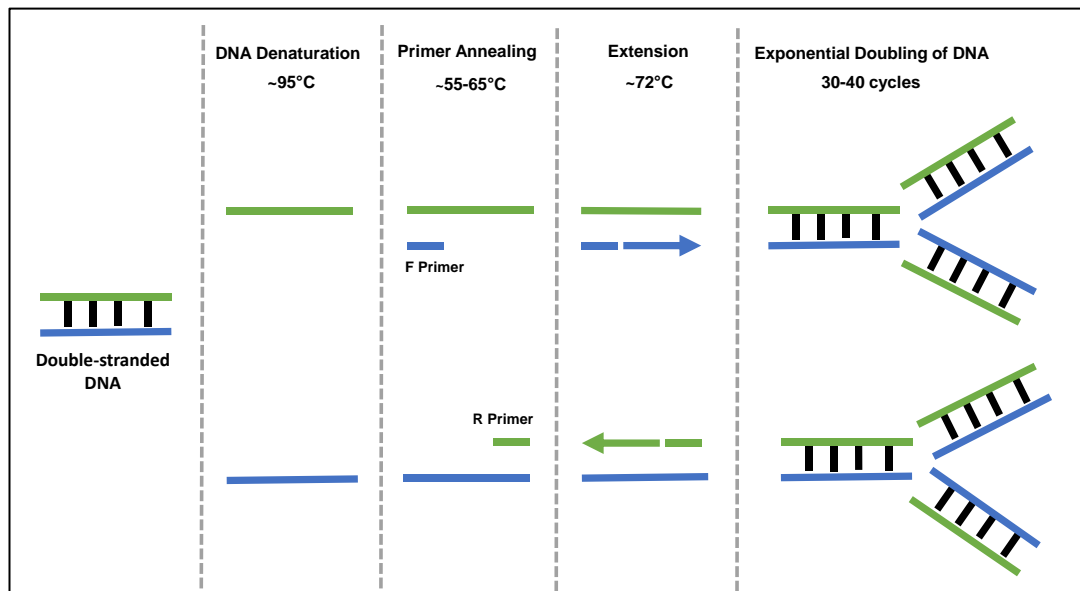
Polymerase Chain Reaction (PCR) is a technique employed to enzymatically replicate nucleic acid and consequently amplify deoxyribonucleic acid (DNA). Significant early research into DNA replication was carried out by Kleppe including a key theoretical description of DNA amplification using primers (Kleppe *et al.*, 1971). The first experimental PCR was carried out in 1985 by Kary Mullis and the technique was advanced significantly with the introduction of a thermostable polymerase isolated from *Thermus aquaticus* (Saiki *et al.*, 1988). This prevented the laborious addition of fresh polymerase following each cycle and improved the quality of the

### Chapter 3.1.2 Polymerase Chain Reaction

final PCR product. Subsequent adaptations of this technology also led to the advancement of DNA sequencing methods.

The reaction comprises *Taq* DNA polymerase, free deoxynucleotide triphosphates (dNTP's), bivalent cations, buffer solution, oligonucleotide primers and a source of DNA template. Forward and reverse primers are designed complementary to a known DNA sequence and so provide the specificity of the reaction. The first step of PCR (Figure 3-1) is denaturation during which the double stranded DNA is melted into two single strands at a high temperature. The temperature is then reduced to promote primer annealing to the DNA template. The final step of the reaction is the extension of the newly synthesised DNA strand by incorporation of free dNTPs by *Taq* polymerase, at an optimum temperature required for polymerase activity ( $\sim 72^{\circ}\text{C}$ ). The reaction repeats for 30-40 cycles until the reaction enters a plateau phase whereby DNA synthesis efficiency is reduced. Possible reasons for the plateau include limiting of primers or dNTP's and accumulation of PCR products shielding polymerase activity.

The goal of PCR optimisation is to determine PCR conditions that maximise specificity and yield of the PCR product. Success of the PCR is usually measured by visualisation of PCR products using DNA gel electrophoresis (yield) and/or DNA sequencing (specificity).



**Figure 3-1 Polymerase Chain Reaction**

Stages of the Polymerase Chain Reaction (PCR) cycle and subsequent doubling of DNA. Each step of the cycle is named and the temperature at which the reaction is carried out.

## 3.2 Method

Methods specific to the development of a PCR assay are listed in the section. All other general methods including PCR, DNA extraction and DNA agarose gel electrophoresis are described in Chapter 2.

### 3.2.1 DNA sequence retrieval and PCR primer design

Careful design of forward and reverse oligonucleotide primers is imperative for PCR specificity. The complementarity of the 3' end of the primer to the target DNA is particularly important as DNA polymerase requires the binding of primers at the 3' end to begin elongation.

DNA sequence alignments using the MUSCLE algorithm were constructed using Molecular evolutionary genetic analysis (MEGA) software, version 7.00 (Kumar *et al.*, 2016). Primers to target bacterial taxa were designed against the 16S rRNA gene. Relevant 16S rRNA sequences were retrieved from the Ribosomal Database Project (RDP). To ensure the accurate sequences for primer design were obtained, only type strain and cultured isolates over 1200 bp were retrieved from the RDP. Primers to target virulence genes were designed using DNA sequences obtained from the genetic sequence databases Genbank at NCBI (<https://www.ncbi.nlm.nih.gov/genbank>) or the European Nucleotide Archive at the European Bioinformatics Institute (<https://www.ebi.ac.uk/ena>).

Metagenomic DNA sequences from clinical samples were retrieved using the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa *et al.*, 2012) basic local alignment search tool (BLAST). Sequences were subsequently added to MEGA DNA sequence alignments for primer design.

Some of the primers used in this study were designed by visual inspection of the constructed DNA sequence alignments. Primer properties required for an efficient PCR are recommended to have certain characteristics (Table 3-1) and so primers were designed accordingly. Multiple primer pairs for each target were designed and tested.

Primer length of 18-30 nucleotides
Melting temperature ( $T_m$ ) of the two primers should be between 65°C and 75°C, and within 5°C of each other.
GC content between 40 and 60%, with the 3' of a primer ending in C or G to promote binding.
A balanced distribution of GC-rich and AT-rich domains
Avoid runs of 4 or more of one base, or dinucleotide repeats (for example, ACCCC or ATATATAT)
Limit self-complementarity between forward and reverse primers to avoid the creation of primer-dimers which may reduce efficiency of the reaction

**Table 3-1 Primer Characteristics recommended for an efficient PCR**

(<https://www.thermofisher.com/uk/en/home/technical-resources/technical-reference-library/pcr-cdna-synthesis-support-center/end-point-pcr-primers-support/end-point-pcr-primers-support-getting-started.html>)

An online multiple primer designer calculator (<https://www.thermofisher.com/uk/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/multiple-primer-analyzer.html>) was used to determine the characteristics of the designed primers.

### **3.2.2 Primer specificity testing**

To select suitable bacterial species for primer specificity testing, phylogenetic trees were constructed. Phylogenetic trees were constructed in MEGA using the distance-based Neighbour-Joining method with maximum composite likelihood substitutions to include: transitions + transversions and gaps/missing data treatment: partial deletion.

### **3.2.3 Annealing Temperature Optimisation**

PCRs were carried using a range of annealing temperatures including a central  $T_a$  chosen by calculating 5°C lower than the primer with the lowest melting temperature ( $T_m$ ) (Table 3-2). A non-specific but closely related DNA template and a no template control were also included for each annealing temperature tested.



**Table 3-2 Primers designed and used for specificity testing**

Target	Primer Name	Sequence (5' to 3')	GC content %	T <sub>m</sub> (°C)	Size (bp)
Universal	27F	AGAGTTTGATCMTGGCTCAG	47.5	56.3	20
	1492R	TACGGYTACCTTGTTACGACTT	43.2	57.5	22
Species <i>B. fragilis</i>	BfF1	GGTGGACTGGTAAGTCAGTTGTG	52.2	62.4	23
	BfF2	ACTGTCAGTCTTGAGTACAGTAGAG	44.0	61.3	25
	BfR1	GCAGACTTCGATCCGAAGCTG	55.0	59.4	20
	BfR2	GTTTCCACATCATTCCACTGC	47.6	57.9	21
<i>B. fragilis</i> toxin ( <i>bft</i> )	BFTF1/BFET-F1	GTTAGTGCCCGATGCAG (Avila-Campos <i>et al.</i> , 2007)	55.6	56.0	18
	BFET-TYPE1	ATTGAACCAGGACATCCC (Avila-Campos <i>et al.</i> , 2007)	50.0	53.7	18
	BFET-TYPE2	CGCTCGGGCAACTAT (Avila-Campos <i>et al.</i> , 2007)	60.0	50.6	15
	BFET-TYPE3	CGTGTGCCATAACCCCA (Avila-Campos <i>et al.</i> , 2007)	58.8	55.2	17
	BFTF2	GAACCTAAAACGGTATATGT (Kato <i>et al.</i> , 2000)	35.0	52.9	20
	BFTR1	CAGCTGGGTTGTAGACATCC	62.6	62.6	20
	BFTR3	GTTGTAGACGACATCCCACTGGC (Kato <i>et al.</i> , 2000)	56.5	64.2	23
<i>B. fragilis</i> ubiquitin ( <i>ubb</i> )	UbbF1	GCTGGACCATAACATTAGAGGT	45.5	58.4	22
	UbbR1	GCTAAAGTTCGTCCATCTTCCA	45.5	58.4	22
Genus <i>Fusobacterium</i>	FusF1	ACTGGACAGATACTGACGC	52.6	56.7	19
	FusF2	GAAAGCGTGGGTAGCAAACAGG	54.5	62.1	22
	FusR1	GTGTAGCCCAGCGTATAA	50.0	53.7	18
<i>Fusobacterium nucleatum</i> group	FNucF1	CGCGTAAAGAAAYTGCCTCACAG	50.0	61.5	23
	FNucF2	CGGTACCAACAGAAGAAGTGACG	52.2	62.4	23
	FNucF3	GTA CTGGAGAGGTAAGCGG	57.9	58.8	19
	FNucR1	TATTCACGCGACATTGCTGATTCTG	48.0	63.0	25
<i>Fusobacterium nucleatum</i> subsp. <i>nucleatum</i>	SubspNF1a	ATTATGATTATAGGGCATCCTAG	34.8	55.3	23
	SubspNF1b	GATTATAGGGCATCCTAG	44.4	51.4	18
	SubspNR1	CCTCCTACTCATCGTAGGC	57.9	58.8	19
	SubspNR2	CCTCACGGCTTTGCAACT	55.6	56.0	18
<i>F. nucleatum</i> <i>FadA</i>	FadAF1	GAAGAAAGAGCACAAGCTGA	45.0	55.2	20
	FadAF2	GATGCAGCAAGTTTAGTAGGTG	45.5	58.4	22
	FadAR1	CWGCTTGAAGTCTTTGAGCTCT	45.5	58.4	22
	FadAR2	GCTTGAAGTCTTTGAGCTCTTTG	43.5	58.9	23

**3.2.4 Limit of Detection**

A single colony of the relevant bacterium was grown for 24 hours in 20 mL pre-reduced BHI. A sample (100 µL) of the culture was used to make serial dilutions and plated onto FAA + 5% horse blood agar plates to determine the number of colony forming units (CFU/mL). The remainder of the sample dilution was centrifuged at 12 000 g for 10 mins and the supernatant

### Chapter 3.3.1 DNA Extraction and Quality

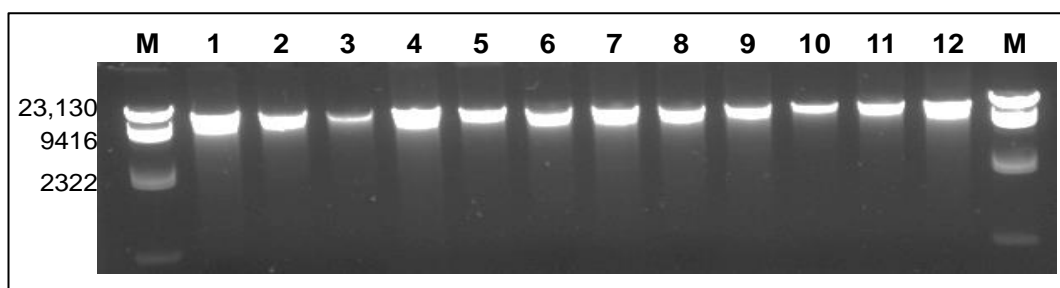
removed. The bacterial pellet was washed by resuspending in an equal volume of sterile water. The bacterial pellet was centrifuged again at 12 000 g for 10 mins and resuspended in an equal volume of sterile water. Dilutions were made from this bacterial suspension and used as DNA templates (1  $\mu$ L) for PCR.

## 3.3 Results

### 3.3.1 *DNA Extraction and Quality*

Good quality DNA template and optimal template concentration are required for a successful and efficient PCR. Quality of DNA extracted from patient samples was assessed by a combination of gel electrophoresis and Nanodrop<sup>TM</sup> spectroscopy. Concentration and quality was initially assessed using Nanodrop spectroscopy using a selection of pilot colonic biopsy samples. The 260/280 nm ratio (considered an indicator of purity) was between 1.84 and 2.00, indicative of the presence of pure DNA without protein and phenol contamination. Subsequent experiments (section 3.3.4) were carried out to test the suitability of the DNA for its downstream application in PCR.

Visualization of the genomic DNA using DNA gel electrophoresis was used to gauge quality via the assessment of DNA smearing as an indicator of sheared genomic DNA. A consistent volume (1  $\mu$ L) of extracted biopsy DNA was loaded for each sample to also give an approximation of DNA quantity. Figure 3-2 shows an example of DNA extracted from biopsy tissue with each lane representing a different sample extract. The gel shows good quality DNA as shown by a discrete band at a large size with limited amounts of shearing.

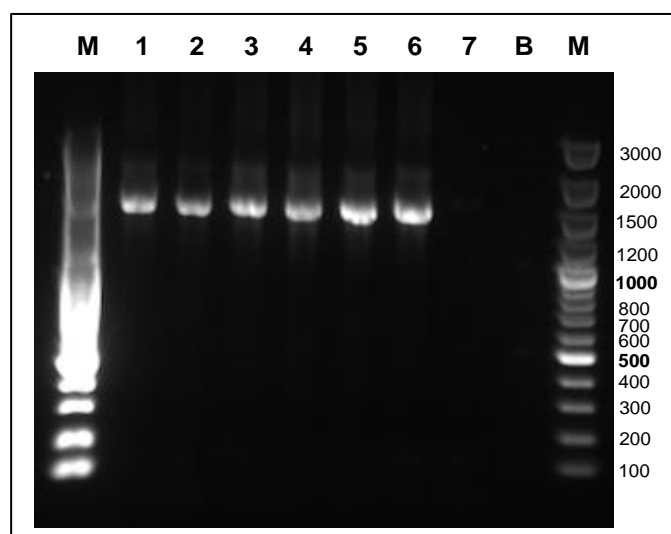


**Figure 3-2 Assessment of DNA Quality from DNA Extracts**

L= DNA ladder ( $\lambda$  DNA-Hind III digest). Marker sizes (bp) are as follows: 23130, 9416, 6557, 4361, 2322 and 2027. Lanes are numbered 1-12 representative of DNA extracts from different colonic biopsy samples.

There was some variability in DNA yield, as determined by fluorometric quantitation, but was typically ~2-10  $\mu$ g (mean = 4.0  $\mu$ g). The DNA extraction procedure was not successful for some samples and so they were excluded from any downstream analysis.

An initial PCR, without further optimisation of conditions, was performed using primers designed to 'universally' amplify all bacterial 16S rRNA DNA templates (Figure 3-3). The resultant PCR products show that the sample DNA extracts did contain PCR-facilitated amplifiable bacterial DNA. The results are unlikely to be from contaminating DNA, as a no template control was negative. Additional experimental testing for the presence of PCR inhibitors was carried out later (see section 3.3.4) following further PCR condition optimization.



**Figure 3-3 Amplification of bacterial 16S rRNA DNA from patient biopsy samples using universal bacterial primers (27F, 1292R)**

(L) = GeneRuler 100bp Plus DNA ladder. Ladder marker sizes (bp) are given to the right of the gel with brighter reference bands labelled in bold. (1)=U193, (2)=U199, (3)=U198, (4)=U205, (5)=U159, (6)=U195, (7)=No DNA template, (B)=Blank lane. Expected size of PCR amplicon = 1500bp.

### 3.3.2 Primer Design and Specificity Testing

The universal primers (27F and 1492R) confirmed that bacterial colony DNA (1  $\mu$ L of a bacterial colony in 50  $\mu$ L sterile water) was amplifiable using the PCR protocol (data not shown). PCR amplicons were submitted for DNA sequencing and an NCBI BLAST search using the sequences confirmed identity of the cultivated isolates to be used for primer specificity testing.

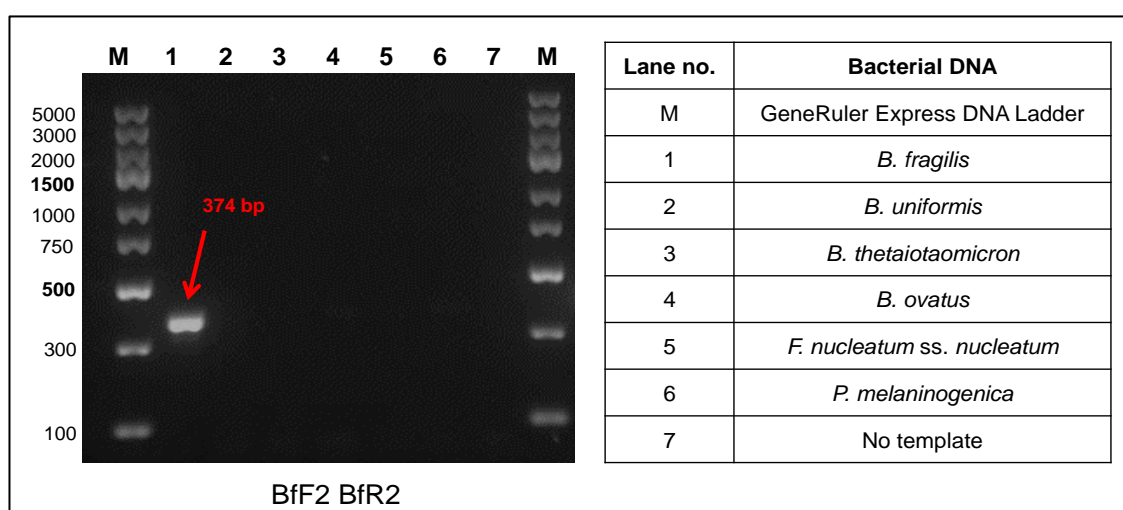
#### 3.3.2.1 *Bacteroides fragilis*

##### Primer design

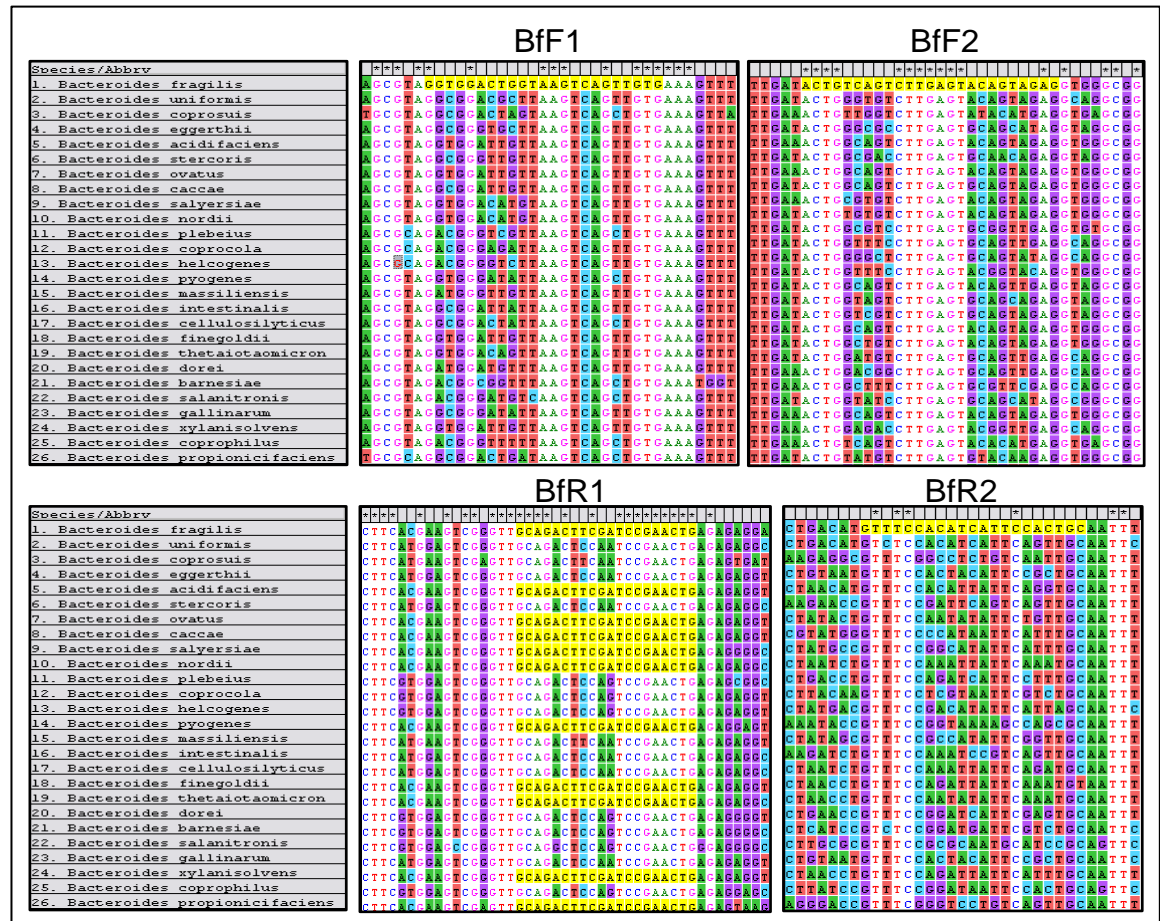
Multiple forward and reverse primers were designed with the aim of being specific for the *B. fragilis* species. Primers for the detection of the bacterial taxa *B. fragilis*, *Fusobacterium* and *F. nucleatum* were designed against the 16S rRNA gene. DNA sequence alignments were created to identify relevant conserved or variable regions of sequence and to determine feasibility of primer design for these targets (Figure 3-4). Sequences were aligned, and primers designed by eye aiming to include as many nucleotide differences as possible in comparison to closely related sequences.

**Primer specificity:**

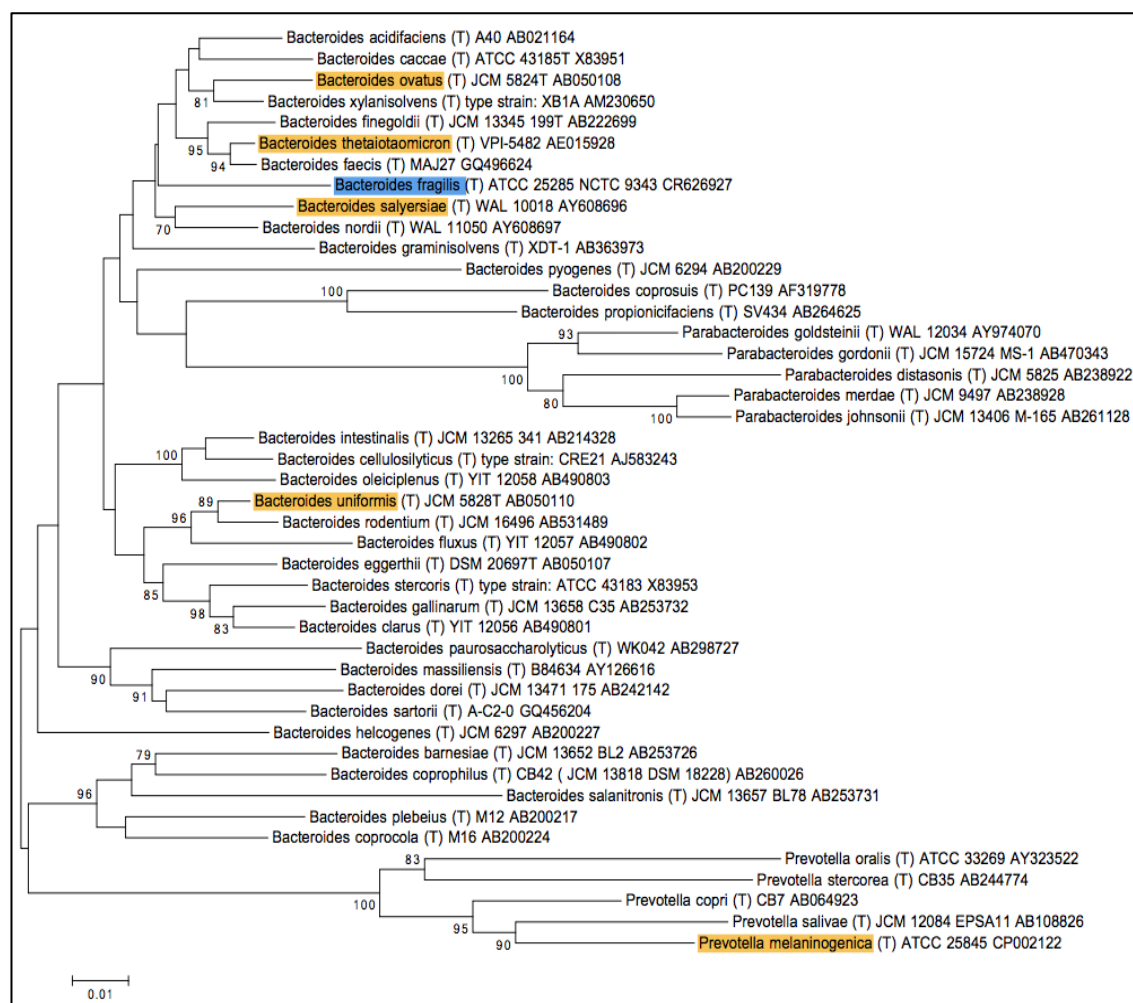
A phylogenetic tree (Figure 3-6) was constructed from the DNA sequence alignment used to design primers (Figure 3-5). *In vitro* specificity testing of *B. fragilis* primer pairs was carried out using a panel of species selected to span a range of *Bacteroides* 16S rRNA sequences and species closely related to *B. fragilis*, as inferred from the phylogenetic tree (Figure 3-6). In addition, a distantly related species was also tested (*F. nucleatum* subsp. *nucleatum*). Specificity testing showed primer pair BfF2 BfR2 to be specific for *B. fragilis*, generating a single amplicon of the correct size (Figure 3-4).

**Figure 3-4 Primer Specificity Testing:**

*B. fragilis* Specificity of *B. fragilis* species-specific primers using genomic DNA from different bacteria. The red arrow highlights the desired amplicon and indicates the band size. Each lane represents a PCR using a different bacterial DNA template as detailed in the adjacent table.

Figure 3-5: Primer Design: *B. fragilis*

DNA sequence alignment of 16S rRNA gene sequences from *B. fragilis* and *Bacteroides* species. The alignment was produced using MEGA 6 software and MUSCLE alignment algorithm. Primers are highlighted in yellow while bases with a white background indicate a conserved base across all sequences in the alignment.



**Figure 3-6 Evolutionary relationships of *Bacteroides* 16S rRNA gene and closely related *Bacteroidales* taxa**

The sequence of interest, *Bacteroides fragilis*, is highlighted in blue while other species used for in vitro specificity testing are highlighted in yellow. The evolutionary history was inferred using the Neighbour-Joining method (Saitou and Nei, 1987). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein, 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura *et al.*, 2004) and are in the units of the number of base substitutions per site. The analysis involved 44 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 1237 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar *et al.*, 2016).

### 3.3.2.2 *Fusobacterium*

**Primer Design:** The *Fusobacterium* DNA alignment (Figure 3-7A) showed three regions of the 16S rRNA gene sequence conserved amongst *Fusobacterium* species. The designed primers are also specific for *Clostridium rectum* and may select for some *Cetobacterium* species. This issue is discussed further in the discussion (section 3.4.2).

**Primer specificity:** Species selected for primer specificity testing are highlighted in yellow in Figure 3-8. All subspecies of *F. nucleatum* and the most distantly related *Fusobacterium*, *F. mortiferum* generated a discrete PCR amplicon of the expected size (Figure 3-7B).

### 3.3.2.3 *Fusobacterium nucleatum*

**Primer Design:** The generated DNA alignment showed few regions of the 16S rRNA gene that were specific for the *F. nucleatum* group and excluded other *Fusobacterium* species (Figure 3-9A).

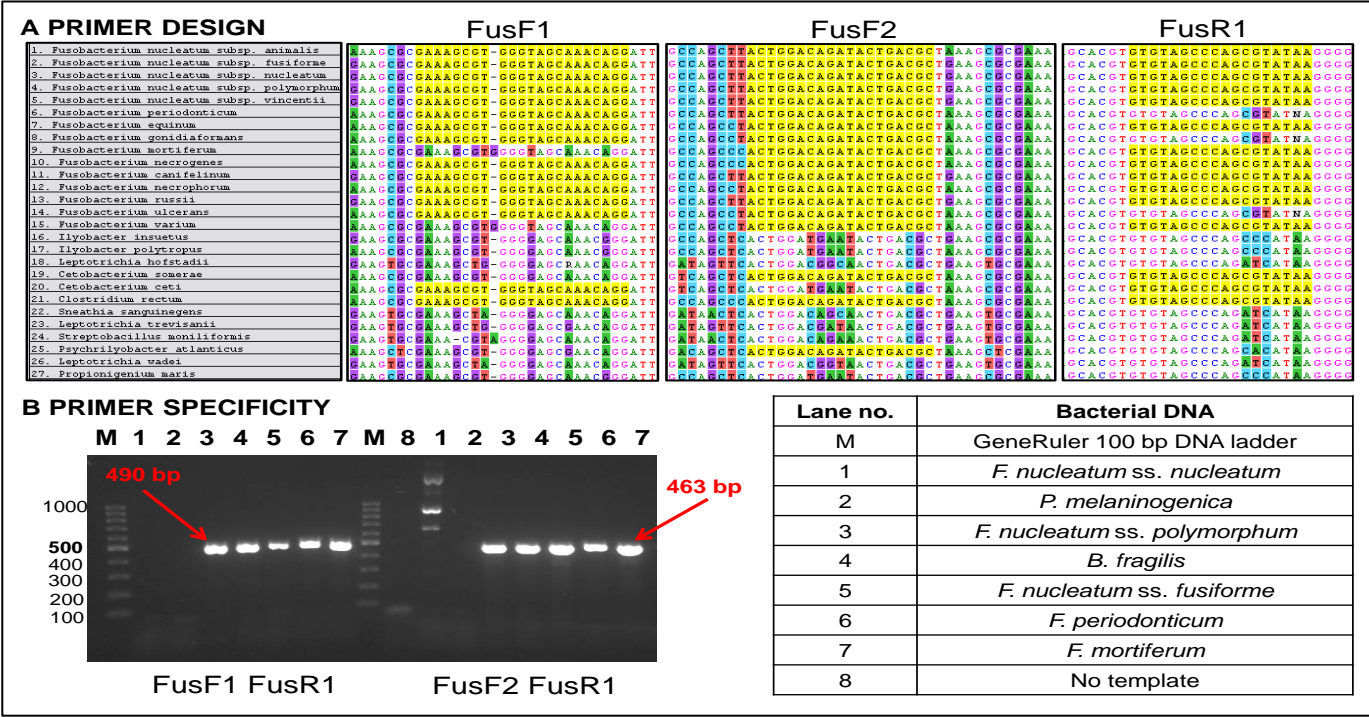
**Primer Specificity:** Specificity testing showed that all primer pair combinations generated a PCR product of the correct size for the most distantly related *Fusobacterium* species (Figure 3-8), *F. mortiferum* (Figure 3-9B). A non-specific amplicon was also generated against the *B. fragilis* DNA template, which would potential reduce the PCR efficiency if *B. fragilis* is also present in the extracted patient DNA. The *F. nucleatum* group were excluded from use in the study due to the lack of specificity of the PCR to the intended target.

### 3.3.2.4 *Fusobacterium nucleatum* subsp. *nucleatum*

**Primer Design:** The DNA sequence alignment revealed limited regions of sequence specific to the different *F. nucleatum* subspecies (Figure 3-10A). A number of primers were designed with the aim of being specific for subspecies *nucleatum*, chosen due to its use in previous studies of the oncogenic potential of *F. nucleatum* (Kostic *et al.*, 2013).

**Primer Specificity:** Amplicons were generated against *F. nucleatum* subsp. *polymorphum* as well as, or instead of *nucleatum* (Figure 3-10B). *F. nucleatum* subsp. *nucleatum* was excluded from use in the study due to the lack of specificity of the PCR to the intended target.

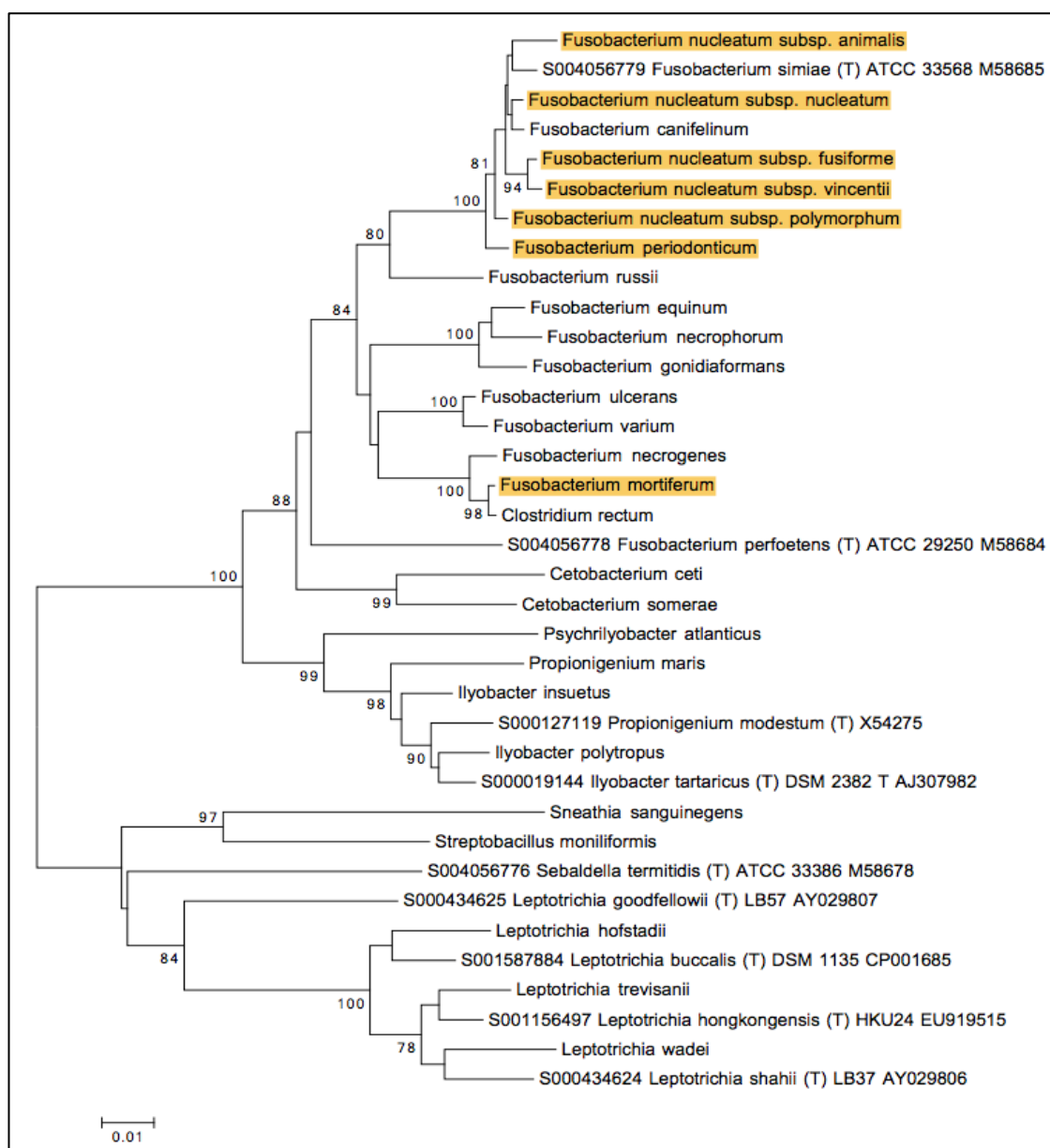




**Figure 3-7 Primer Design and Specificity: *Fusobacterium* genus**

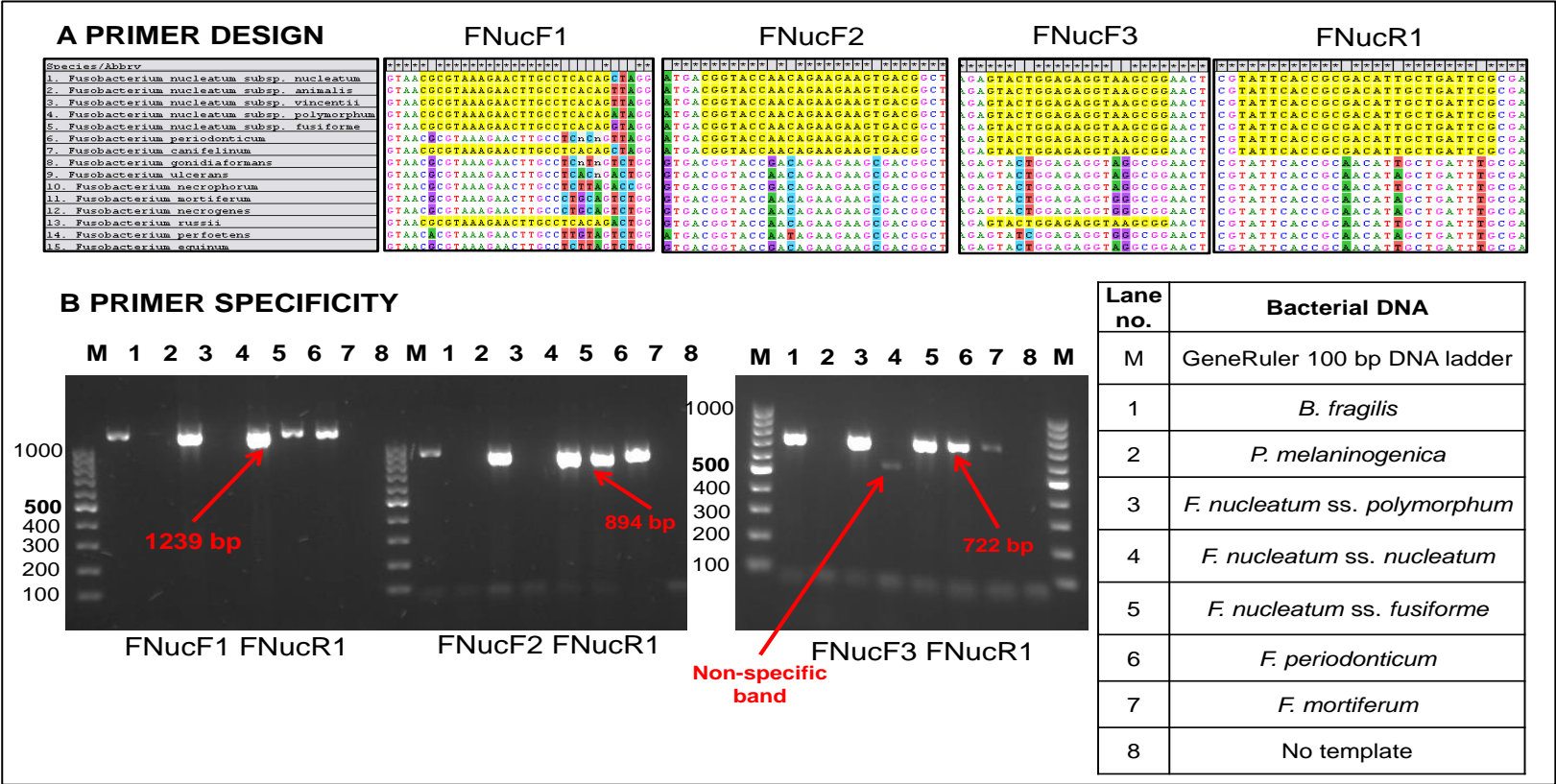
(A) 16S rRNA gene sequences from species in the genus *Fusobacterium* and other species taxonomically related. Letters with a white background indicate a conserved base across all sequences in the alignment. Primers are represented by yellow highlighted bases and are named above. DNA sequence alignment produced using MEGA 6 software and MUSCLE alignment algorithm.

(B) Specificity of *Fusobacterium* genus specific primers using genomic DNA from different bacteria. The red arrow highlights the desired amplicon and indicates the band size. Each lane represents a PCR using a different bacterial DNA template as detailed in the adjacent table.



**Figure 3-8 Evolutionary relationships of *Fusobacterium* 16S rRNA gene and closely related taxa**

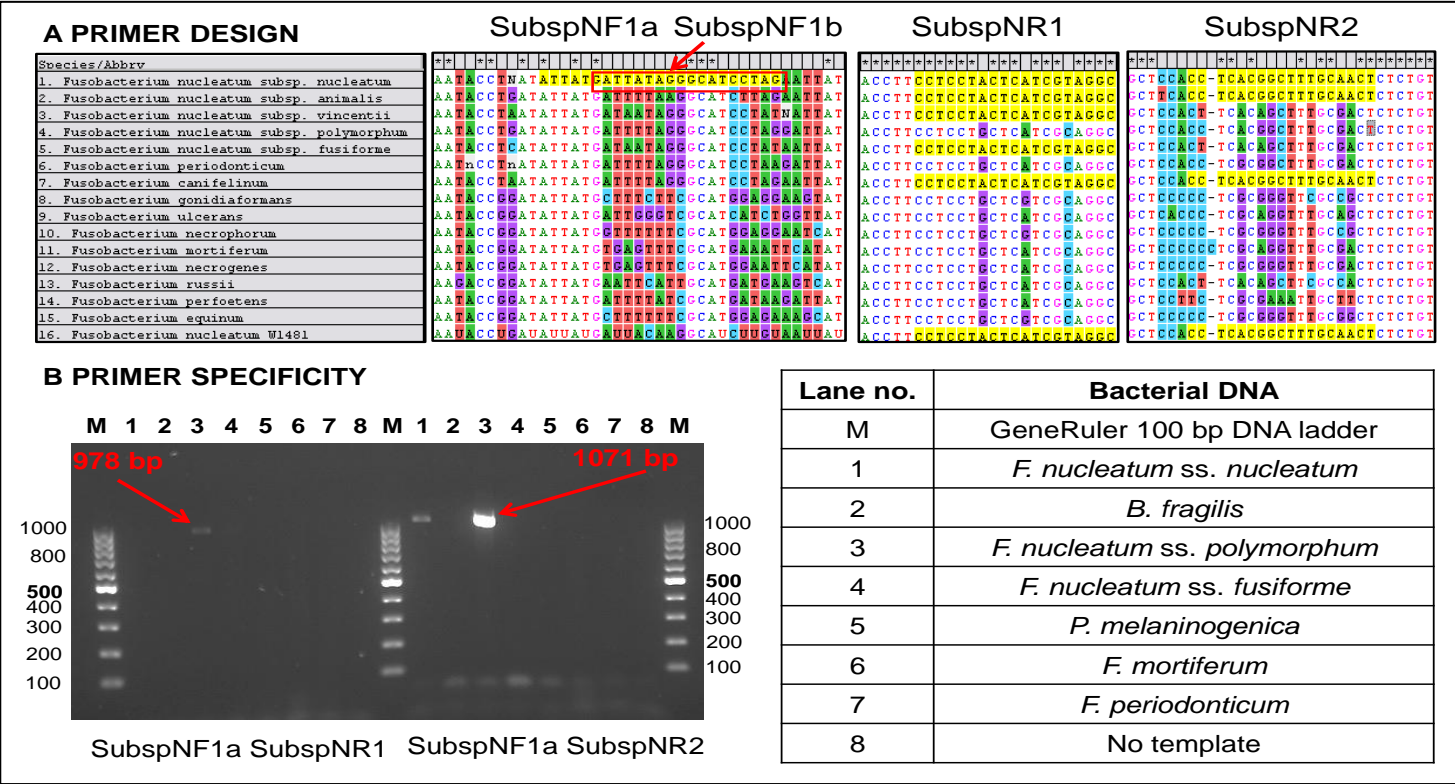
The evolutionary history was inferred using the Neighbour-Joining method (Saitou and Nei, 1987). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein, 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura *et al.*, 2004) and are in the units of the number of base substitutions per site. The analysis involved 44 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 1237 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar *et al.*, 2016).



**Figure 3-9 Primer Design and Specificity Testing: *F. nucleatum* group**

(A) DNA sequence alignment produced using MEGA 6 software and the MUSCLE alignment algorithm. 16S rRNA gene sequences were from species in the genus *Fusobacterium*. Letters with a white background indicate a conserved base across all sequences in the alignment while primers are represented by yellow highlighted bases and are named above.

(B) Specificity of *F. nucleatum* group primers using genomic DNA from different bacteria. The red arrow highlights the desired amplicon and indicates the band size. Each lane represents a PCR using a different bacterial DNA template as detailed in the adjacent table.



**Figure 3-10 Primer Design and Specificity Testing: *F. nucleatum* subspecies *nucleatum***

(A) 16S rRNA gene sequences from species in the genus *Fusobacterium* species and other species taxonomically related. Letters with a white background indicate a conserved base across all sequences in the alignment. Primers are represented by yellow highlighted bases and are named above. DNA sequence alignment produced using MEGA 6 software and MUSCLE alignment algorithm.

(B) Specificity of *F. nucleatum* subspecies *nucleatum* primers using genomic DNA from different bacteria. The red arrow highlights the desired amplicon and indicates the band size. Each lane represents a PCR using a different bacterial DNA template as detailed in the adjacent table.

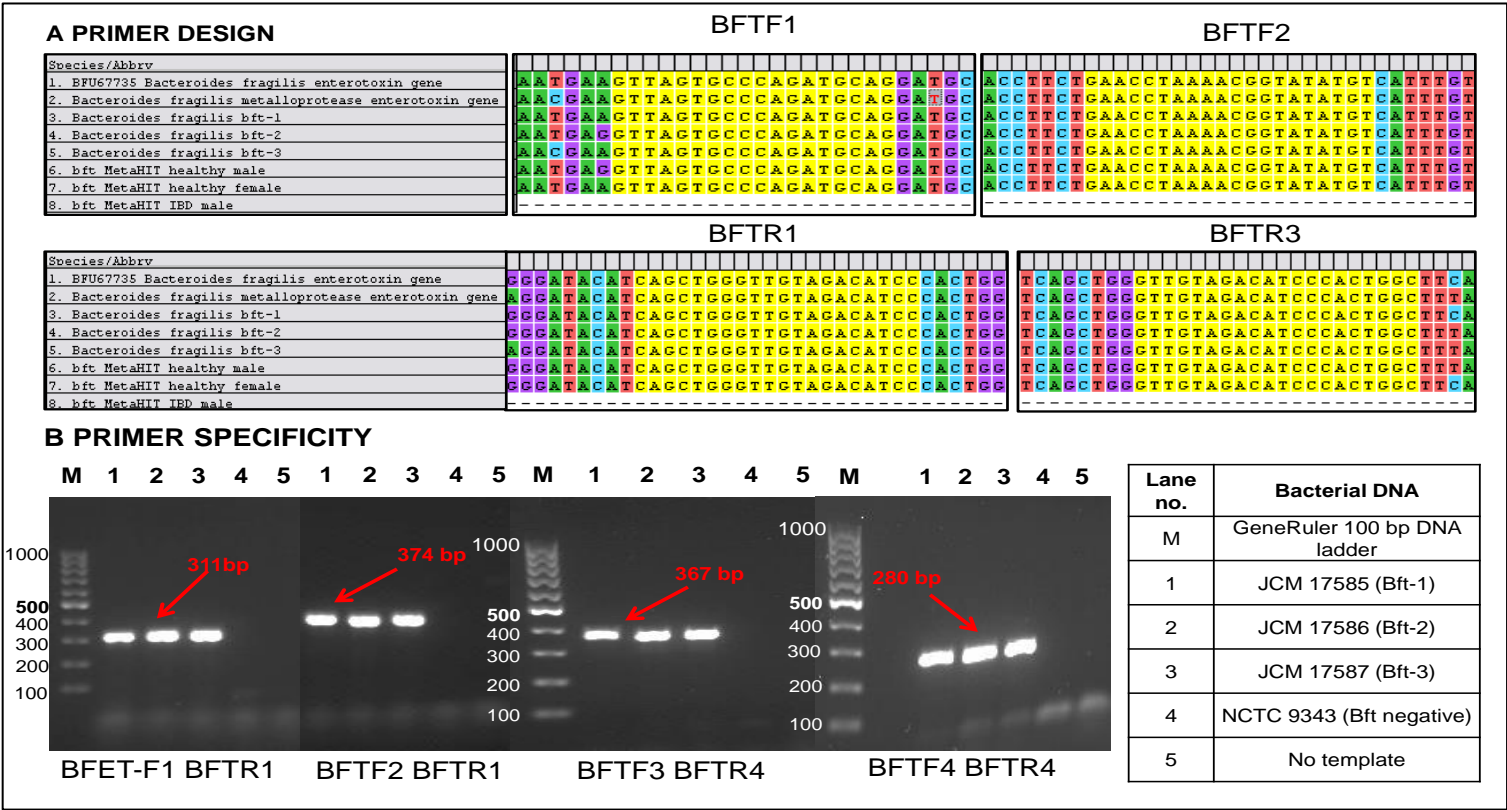
### 3.3.2.5 BFT

**Primer specificity:** *BFT* primers were tested against a panel of *B. fragilis* strains including all three toxin subtypes and the non-toxigenic type strain (Figure 3-11A). A combination of previously published, newly designed, and adapted versions of previous primers were used in specificity testing. PCR amplicons sent for sequencing confirmed the three BFT positive strains were the three different subtypes and therefore could subsequently be used as positive controls.

### 3.3.2.6 *Ubb*

**Primer design:** A DNA sequence alignment for *ubb* primer design was composed of relevant sequences obtained from GenBank (Benson *et al.*, 2013). Additional sequences representative of those found in the gastrointestinal tract were obtained from MetaHIT (Qin *et al.*, 2010) stool metagenomic sequences. As all *ubb* sequences were homologous, primers were designed using the online primer design tool Primer3 web v.4 (<http://primer3.ut.ee>). As this gene product was originally of interest due to its homology to human ubiquitin DNA, sequence alignments were modified to include additional similar sequences. The primers designed using the online design tool were compared against these analogous, non-specific sequences (Figure 3-12A)

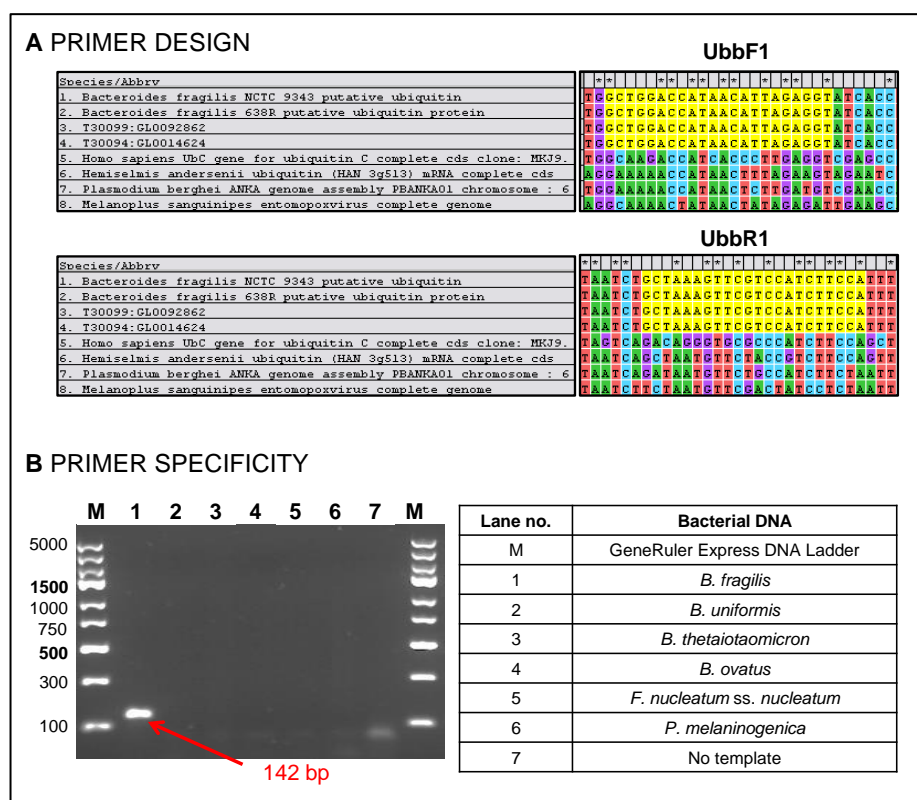
**Primer specificity:** Primers designed to amplify *ubb* were tested against a positive control (*B. fragilis* NCTC 9343) that is known to contain the gene (Patrick *et al.*, 2011). There are no other bacterial species that are known to contain ubiquitin-like genes therefore the specificity of the primers was tested against the previously used panel of bacteria. One discrete amplicon of expected size was generated using the positive control DNA template (Figure 3-12B) and no PCR products were detected for other DNA templates.



**Figure 3-11 Primer Design and Specificity Testing: *B. fragilis* toxin gene (*bft*)**

(A) DNA gene sequences for the *B. fragilis* toxin gene were obtained from Genbank and MetaHit metagenomic sequences. Letters with a white background indicate a conserved base across all sequences in the alignment, while primers are represented by yellow highlighted bases and are named above. DNA sequence alignment produced using MEGA 6 software and MUSCLE alignment algorithm.

(B) Determination of the specificity of *B. fragilis* toxin primers using genomic DNA from different bacteria. The red arrow highlights the desired amplicon and indicates the band size. Each lane represents a PCR using a different bacterial DNA template as detailed in the adjacent table.



**Figure 3-12 Primer Design and Specificity Testing: *B. fragilis* ubiquitin gene (*ubb*)**

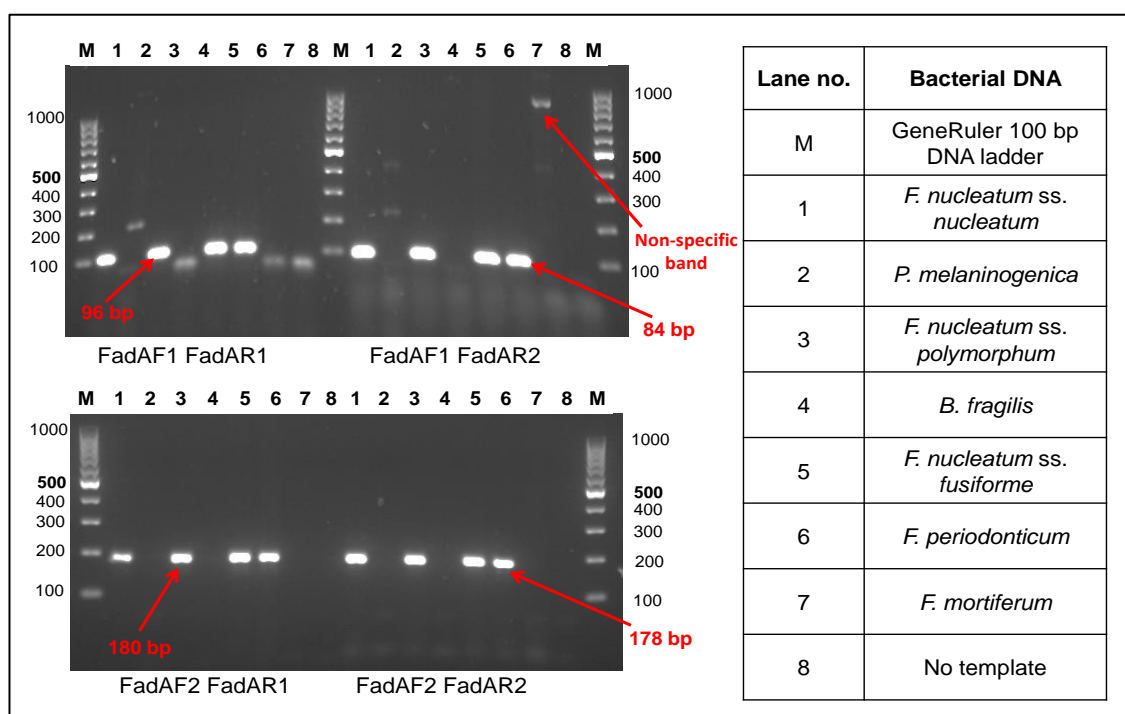
(A) DNA gene sequences for the *B. fragilis* ubiquitin gene were obtained from Genbank and MetaHit metagenomic sequences. Letters with a white background indicate a conserved base across all sequences in the alignment, while primers are represented by yellow highlighted bases and are named above. DNA sequence alignment produced using MEGA 6 software and MUSCLE alignment algorithm.

(B) Specificity of *B. fragilis* ubiquitin-like protein primers using genomic DNA from different bacteria. The red arrow highlights the desired amplicon and indicates the band size. Each lane represents a PCR using a different bacterial DNA template as detailed in the adjacent table.



**3.3.2.7 *fadA***

Primers targeting *fadA* were adapted from previously designed primers (Avila-Campos *et al.*, 2007, Rubinstein *et al.*, 2013) so DNA alignments were not constructed. Primer pairs FadA F1R1 and FadA F1R2 both produced non-specific amplicons to non-specific bacterial templates. Both FadA F2R1 and FadA F2R2 primer pairs were shown to be suitable for the PCR assay following primer specificity testing (Figure 3-13).



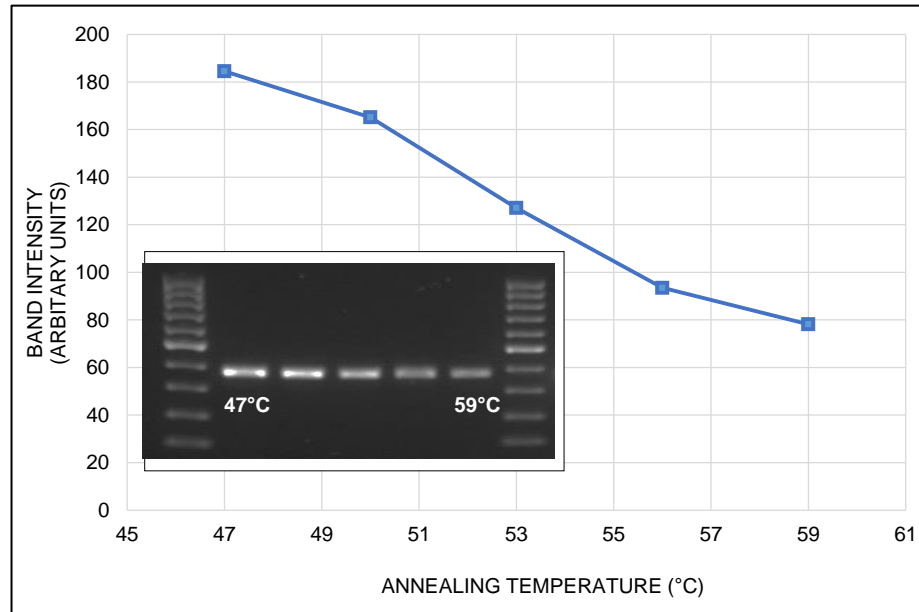
**Figure 3-13 Primer Specificity Testing: *Fusobacterium fadA***

Specificity of *Fusobacterium fadA* primers using genomic DNA from different bacteria. The red arrow highlights the desired amplicon and indicates the band size. Each lane represents a PCR using a different bacterial DNA template as detailed in the adjacent table.

**3.3.3 Annealing Temperature Optimisation**

A summary of the optimal annealing temperature for each primer pair are listed in the table of optimised PCR conditions (Table 3-3). Figure 3-14 represents an example experiment to determine an optimal annealing temperature. At lower annealing temperatures the fidelity of primer binding is decreased therefore an increased PCR product is generated but with a reduced specificity. Primer pairs were tested over a range of annealing temperatures to find the optimum temperature whereby the maximum desired product was generated with minimal non-specific amplification products (McPherson and Moller, 2006).





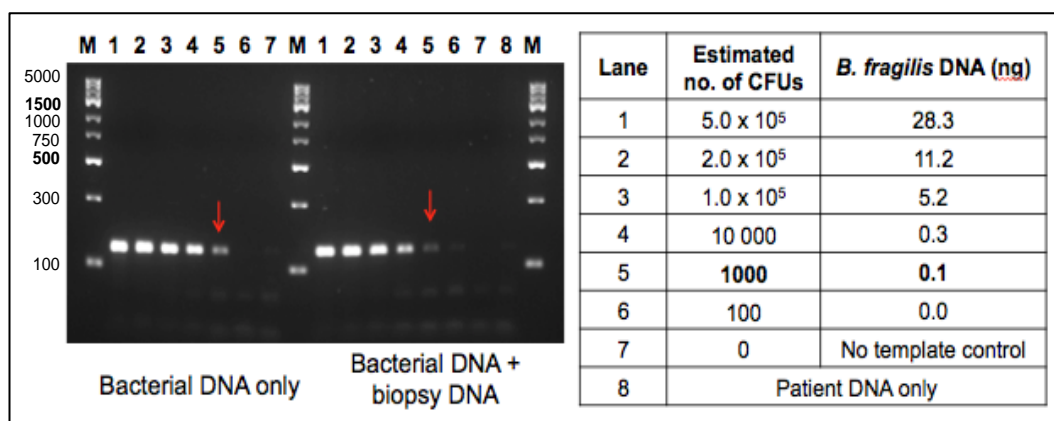
**Figure 3-14 Annealing temperature optimisation: *B. fragilis***

Graph showing the amplicon band intensity using *B. fragilis* species-specific primers (BfF2 BfR2) across different annealing temperatures (47°C, 50°C, 53°C, 56°C, 59°C). The insert contains a representative DNA agarose gel with amplicons generated across different annealing temperatures.

### 3.3.4 PCR Inhibitors and Limit of Detection

PCR using dilutions of *B. fragilis* DNA only and *B. fragilis* DNA spiked into extracted biopsy DNA were performed in parallel to assess differences in sensitivity of the reaction. Figure 3-15 shows comparable amplicon intensities across DNA template dilutions, using both bacterial DNA only and bacterial DNA spiked into extracted biopsy DNA (40 ng) thought to be negative for *B. fragilis*. These results indicate that the extracted biopsy DNA was not inhibiting the PCR reaction.

Due to the absence of PCR inhibitors the limit of detection was subsequently approximated for each gene target using serial dilutions of the relevant bacterial DNA. Figure 3-15 shows that lane 5 is the lowest concentration of DNA producing an amplicon, as highlighted by the red arrow, and in this case represents 1000 colony forming units (CFUs). As for *ubb* all other gene targets (data not shown) had a limit of detection below 1000 CFUs or 0.1 ng *B. fragilis* DNA.



**Figure 3-15 Determination of PCR inhibition using spiked biopsy DNA: *ubb***

Dilutions of *B. fragilis* DNA, listed in the adjacent table, were used in PCR reactions for *ubb*. The equivalent dilutions of *B. fragilis* DNA were added to 40ng extracted biopsy DNA thought to be negative for *B. fragilis*. The red arrows indicate the limit of detection.

### 3.3.5 Results Summary

A summary of the optimized PCR conditions that were used to investigate patient samples are detailed in Table 3-3. The results of these PCRs used to compare patient cohorts are described in chapter 4.

**Table 3-3 Primers and optimised PCR Conditions used for**

Primer Target	Forward Primer	Reverse Primer	PCR Conditions	Expected Size PCR Product
Universal	27F	1492R	30 cycles; Ta = 52 °C; Ext = 90s	1465 bp
Genus <i>Fusobacterium</i>	FusF1	FusR1	30 cycles; Ta = 55°C; Ext = 45s	490 bp
<i>fadA</i>	FadAF2	FadAR1	30 cycles; Ta = 59°C; Ext = 25s	180 bp
<i>B. fragilis</i> species	BfF2	BfR2	30 cycles; Ta = 59°C; Ext = 45s	374 bp
<i>bft</i>	BFTF2	BFTR1	30 cycles; Ta = 48°C; Ext = 25s	311 bp
<i>ubb</i>	UbbF1	UbbR2	30 cycles; Ta = 56°C; Ext = 25s	142 bp

Ta = Annealing temperature, Ext = Extension time

### 3.4 Discussion

The results presented in this chapter determined the primers and PCR conditions to be used for the analysis of patient samples. Based on these results it was concluded that the optimised PCRs to probe patient samples would target *B. fragilis*, *bft*, *ubb*, the genus *Fusobacterium* and *fadA*.

#### 3.4.1 DNA template

A successful DNA extraction requires complete cell lysis, removal of contaminants and minimal shearing of the DNA. Microbial communities have a diverse range of cells walls and consequently it is essential that DNA extraction procedures allow complete cell lysis, particularly for bacteria with thick peptidoglycan layers. Equally a DNA extraction procedure cannot be too harsh that it comprises DNA quality.

Biopsy tissue contains many possible inhibitors of the PCR reaction, including blood components, bile and complex polysaccharides (Schrader *et al.*, 2012). Common contaminants remaining from the DNA extraction procedure that may also inhibit PCR are guanidine HCl and chaotropic salts. In the presence of PCR inhibitors lower amounts of DNA template will result in more PCR product as the inhibitors are diluted out (Wiedbrauk *et al.*, 1995).

In this study DNA quality, as assessed by Nanodrop™ spectroscopy and DNA gel electrophoresis showed limited shearing and degradation by DNases (Figure 3-13) and limited presence of contaminants (Figure 3-15). Further analysis of the extracted DNA for its downstream application of PCR demonstrated a negligible effect of any PCR inhibitors (Figure 3-15) in the sample and amplifiable bacterial DNA (Figure 3-14). Blank controls added to DNA extractions (data not shown) had below detectable limits of DNA, as determined by Qubit™ fluorometer and universal 16S rRNA PCR, suggesting the method of DNA extraction did not add any (or below detection) exogenous DNA either from the DNA kit or the surrounding environment.

In conclusion the DNA extracted from patient biopsy tissue samples was of a satisfactory quantity and quality to investigate the prevalence of the determined gene targets using PCR.

### 3.4.2 Detection methods

Methods of investigating the CRC microbiome by sequencing 16S rRNA do not give sufficient resolution to investigate the role of *B. fragilis* in colonic tumour development. Only select *B. fragilis* strains contain the *bft* gene and therefore either metagenomic sequencing or a method specifically targeting *bft* is required.

A study comparing standard, quantitative and digital PCR methods as ways of detecting *bft* found standard end-point PCR to be the least sensitive (Purcell *et al.*, 2016). The development of an end-point PCR assay described in this thesis used different *bft* primers and a series of optimisation steps aimed at maximising sensitivity. Another difference is that Purcell *et al.*, 2016 used faecal samples compared with colonic biopsy samples used in this study.

Other studies investigating *bft* prevalence in patient samples used culture-based methods to determine *bft*-positivity (Toprak *et al.*, 2006, Boleij *et al.*, 2015). An advantage of using DNA extracted from biopsies frozen on collection is that the samples provide a snapshot in time of the microbiome. The results described in this chapter show the design of primers that capture all *in vivo* *bft* sequences, including all three known subtypes. To ensure the specificity of the primers, sequences were retrieved from previously published metagenomic studies; this was particularly important given that it has been acknowledged (Aitchison *et al.*, 2016) that a previously published sequence (Moncrief *et al.*, 1998) was incorrect. Aitchison *et al.*, 2016 compared *bft* primers that have been used in previous literature investigating *bft* prevalence and found that these primer sets do not detect all *bft* subtypes.

An advantage of using a culture-independent method of *bft*-detection is that bacteria have mechanisms to outcompete other species or even strains from the same species during culturing. *B. fragilis* uses type VI secretion systems (T6SS) to target *Bacteroidales* strains (Chatzidaki-Livanis *et al.*, 2016) and a non-toxigenic strain has been shown to outcompete ETBF using a T6SS (Hecht *et al.*, 2016). These competition mechanisms may result in an altered outcome when using culturing-dependent methods for bacterial identification.

The results presented in this chapter showed that using the 16S rRNA gene to specifically target either *F. nucleatum* as a group or individual subspecies was not possible (section 3.3.2). In

order to specifically target *F. nucleatum*/individual subspecies alternatives to the 16S rRNA gene for bacterial identification could be investigated, such as the beta subunit of DNA polymerase *rpoB*. *fadA* was considered a better target for the *F. nucleatum* group than using the 16S rRNA gene, especially due to the association of *fadA* with virulence (Rubinstein *et al.*, 2013). Primer pairs targeting the genus *Fusobacterium* were also designed, as species aside from *F. nucleatum* have been associated with CRC and IBD (Ohkusa *et al.*, 2003, Drewes *et al.*, 2017). The primers described in this thesis are likely to also detect *Clostridium rectum*, however *C. rectum* is phylogenetically closely related to *F. mortiferum* (Figure 3-8), and despite phenotypic differences should arguably be reclassified within the *Fusobacterium* genus (Lee *et al.*, 2016a). Future work would further analyse the specificity of the *Fusobacterium* primers, particular against *Cetobacterium* species, in order to make conclusions based on the prevalence results described in section 4.2.4.

Future work would ideally use RT-PCR to quantify and compare expression of *bft* and *mpll* (section 1.7.5). MPIL can bind to, and subsequently prevent cleavage, of the adherens junction protein E-cadherin (Remacle *et al.*, 2014) suggesting BFT and MPIL have distinct functions. The literature suggests these two metalloproteinases are counter transcribed (Remacle *et al.*, 2014) and therefore this future work would be a step towards identifying their *in vivo* expression conditions.

## Chapter 4 : PREVALENCE OF BACTERIA AND GENES OF INTEREST IN PATIENT SAMPLES

### 4.1 Introduction

Several metagenomic sequencing studies have revealed a different microbiome in patients with CRC compared to healthy individuals and an altered microbiome on the tumour tissue compared with adjacent healthy tissue (Marchesi *et al.*, 2011).

### 4.2 Results

#### 4.2.1 Clinical results

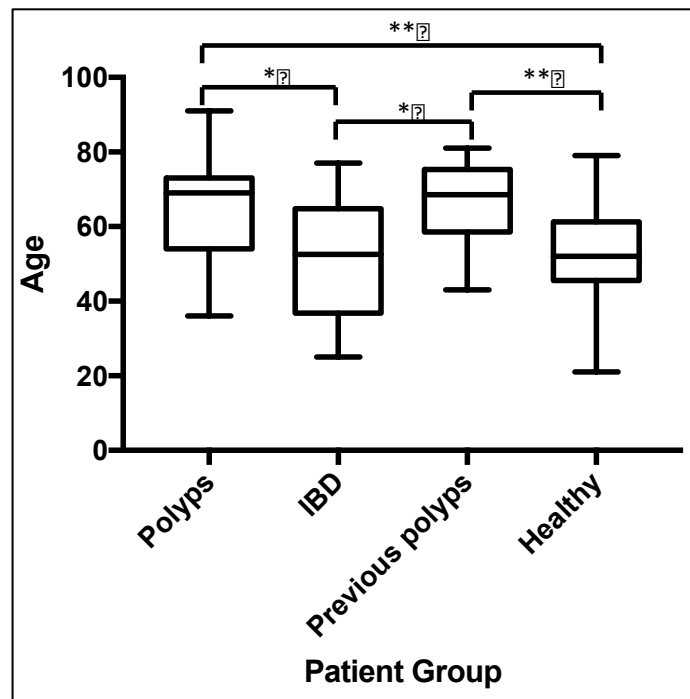
A total of 78 individuals were recruited for the study (Table 4-1) however some IBD patients were also included in other categories (either polyp/cancer or previous polyps/cancer) therefore giving a total of 85. Biopsy samples were collected from all individuals while caecal effluent was collected from 59 patients and blood from 60. Individuals were included in the previous polyps/cancer group if the current colonoscopy was all clear, if new polyps were found they were included in the newly diagnosed polyp/cancer cohort. A separate analysis of individuals with a history of polyps/cancer and returning or newly diagnosed polyps is seen in section 4.2.6.

**Table 4-1 Summary of individuals and samples recruited for the study**

Reason for attendance for colonoscopy	Number of individuals recruited	Number of individuals with caecal effluent samples collected	Number of individuals with blood samples collected
Newly diagnosed polyp/cancer	33	18	21
Inflammatory Bowel Disease (IBD)	15	10	11
Previous polyps/cancer	14	10	10
Healthy (anaemia or rectal bleeding)	23	21	18
Total	85	59	60

Figure 4-1 shows that there was a significant difference in age between some of the patient cohorts. The newly diagnosed polyp/cancer cohort had a mean age of 64, the IBD group 51, the previous polyps/cancer group 66, and the healthy group 51. The IBD and healthy group had a significantly lower mean age than the polyps/cancer ( $p$  values = 0.021 and 0.007 respectively)

and previous polyps/cancer (p values = 0.011 and 0.004 respectively) and so interpretation of subsequent sample analysis should take this into account.



**Figure 4-1 Age of individuals recruited to each patient group**

A box plot showing the age distribution amongst patient groups. The box represents the interquartile range of the data with the horizontal line signifying the median. The range of the data is represented by the 'whiskers' of the box. \* =  $p < 0.05$ ; \*\* =  $p < 0.01$

A Chi-square statistical analysis showed no significant difference ( $p = 0.727$ ) between the numbers of male or females individuals recruited into the different patient cohorts.

#### 4.2.2 Extracted DNA

A one-way ANOVA test showed no significant difference in total DNA extracted from biopsy samples between patient groups ( $p = 0.151$ ). There was some variability in the amount of DNA extracted from biopsy samples, but in general approximately 4-7  $\mu\text{g}$  of total DNA was extracted per biopsy sample. DNA extraction was not successful for some samples or the DNA obtained could not be amplified in a universal bacterial PCR using primers 27F and 1492R (6 samples from 4 individuals), and so were excluded from the study. Control DNA extractions using no sample had a DNA concentration below that detectable by the Qubit fluorometer.

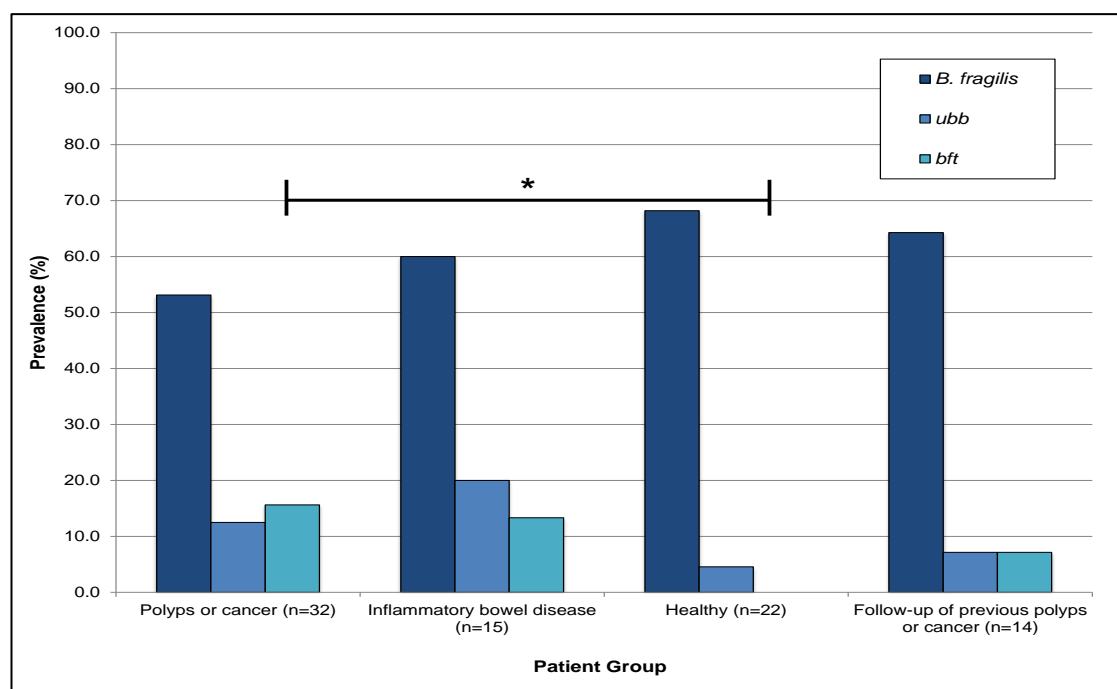
**4.2.3 Prevalence of *B. fragilis*, *bft* and *ubb***

All PCR results for each sample and individual can be found in Appendix D. *B. fragilis* was found in approximately two thirds of individuals, with no significant difference amongst patient cohorts (Table 4-2). *B. fragilis* strains positive for either the *bft* or *ubb* gene were a lot less prevalent with a detection rate of 11%. There was no significant difference in the presence of *ubb* amongst patient groups however there was a significantly higher detection rate of *bft* in patients with newly diagnosed polyps/cancer compared with healthy individuals ( $p = 0.045$ ) (Figure 4-2). Tables 4-3 and 4-4 list the individuals positive for *bft* and *ubb* respectively, including relevant clinical data.



**Table 4-2 Numbers of individuals positive for *B. fragilis*, *bft* and *ubb***

Patient Cohorts	Cases	<i>B. fragilis</i>		<i>bft</i>		<i>ubb</i>	
		Number Detected	Detection Rate (%)	Number Detected	Detection Rate (%)	Number Detected	Detection Rate (%)
Newly diagnosed polyp/cancer	32	17	53	5	16	4	13
Inflammatory Bowel Disease	15	9	60	2	13	3	20
Healthy (anaemia or rectal bleeding)	22	15	68	0	0	1	5
Previous polyps/cancer	14	9	64	1	7	1	7
<b>Total</b>	<b>83</b>	<b>50</b>	<b>60</b>	<b>8</b>	<b>11</b>	<b>9</b>	<b>11</b>

**Figure 4-2 Percentage of individuals (prevalence) PCR-positive for *B. fragilis*, *bft* and *ubb* in biopsy samples**

PCR positivity was defined as one or more biopsy samples with a PCR amplicon for the specified gene target. Statistical significance was determined by Fisher's exact test with a \* designating a p value less than 0.05.

**Table 4-3 Clinical characteristics of *ubb* positive individuals**

Patient ID	<i>ubb</i> location	Patient Group	Age at collection	Sex	Disease status	Medical History
BF003	AC	IBD (Colonic Crohn's)	42	F	Mild patchy inflammation with focal active inflammation	Angina, hypertension
BF013	AC	New polyps	N/A	M	Polyp histology unknown	Previous polyps in AC and caecum
BF018	AC, DC	Healthy	49	M	Rectal bleeding	
BF019	NT	IBD and new polyps	61	M	2 x tubulovillous polyps with low grade dysplasia. 1 in hepatic flexure, 1 in sigmoid colon (~2 cm). Biopsies taken from sigmoid polyp.	
BF033	AC, DC	IBD (Unclassified UC)	34	M	No inflammation	
BF059	AC, DC	Previous polyps	51	F	Previous pedunculated distal, sigmoid polyp	
BF061	OT, NT, AT	New cancer	71	M	Presented with rectal bleeding and constipation. Rectosigmoid cancer	Osteoarthritis, thyroid disease
BF090	AC, DC	New polyp	73	F	1cm rectal tubular adenoma	Diverticular disease, hypertension

CD = Crohn's Disease; UC = Ulcerative colitis; AC = Ascending colon; DC = Descending colon; OT = On tumour; NT = Next to tumour; AT = Away from tumour

**Table 4-4 Clinical characteristics of *bft* positive individuals**

Patient ID	<i>bft</i> location	Patient Group	Polyp location	Age at collection	Sex	Disease status and biopsy histopathology	Other relevant medical history
BF003	AC, DC	IBD (CD)	N/A	42	F	Diagnostic biopsies showed mild patchy inflammation with focal active inflammation	
BF008	OT, NT and AT	Polyp (Cancer)	Proximal AC	71	F	Transverse serrated sessile polyp and 2cm sessile lesion (biopsies taken from this tumour) is adenocarcinoma T3(b) No nodes Dukes B	Recent breast cancer diagnosis. Previous skin cancer
BF027	OT, NT, AT	Polyp	DC	36	M	Pedunculated inflammatory pseudopolyp	
BF028	OT, NT, AT	Polyp	Sigmoid	70	F	Multiple small sessile polyps. Previous tubular adenoma and hyperplastic polyps	
BF033	AC, DC	IBD	N/A	34	M	Unclassified IBD. Azathiopurine treatment for IBD. Diagnostic biopsies showed no inflammation.	
BF034	OT, NT, AT	Polyp	Sigmoid	54	F	Pedunculated tubular adenoma, low-grade dysplasia.	Portal vein thrombosis and varices. HIV Positive
BF043	AC, DC	Polyp	Sigmoid	74	M	Inflammatory polyp	Anaemia. HIV positive.
BF080	AC, DC	Previous polyp	Unknown	75	M	Family history of CRC (mother and father).	Diverticular disease, COPD, gout and hypertension.

PSC = Primary Sclerosing Cholangitis; HIV = Human Immunodeficiency Virus; COPD = Chronic Obstructive Pulmonary Disease; CD = Crohn's Disease; UC = Ulcerative colitis; AC = Ascending colon; DC = Descending colon; OT = On tumour; NT = Next to tumour; AT = Away from tumour

**4.2.4 Prevalence of *Fusobacterium* and *fadA***

A higher prevalence, although not significant ( $p = 0.152$ ), of the *Fusobacterium* genus was found in the healthy cohort as compared with individuals with newly diagnosed polyps or cancer (Figure 4-3). There was no significant difference amongst patient groups for the *fadA* gene, although across all cohorts *fadA* had a low detection rate (9%) (Table 4-5).

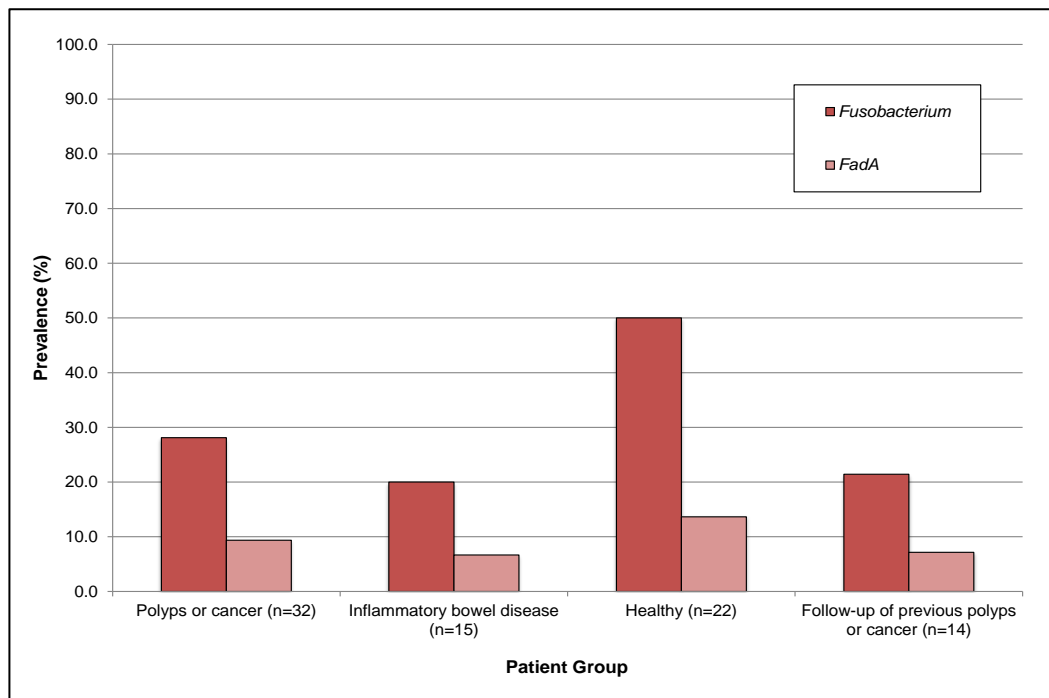
*Fusobacterium* was more likely to be found in a single biopsy location as compared with the other gene targets (Appendix 7.1). Ascending colon biopsies appeared to be more likely to be positive for *Fusobacterium* than descending colon biopsies with a 28% versus 17% detection rate although this was not a significant difference ( $p = 0.186$ ), as determined by a Fisher's exact test.

**Table 4-5 Numbers of individuals positive for *Fusobacterium* and *fadA***

Patient Cohorts	Cases	<i>Fusobacterium</i>		<i>FadA</i>	
		Number Detected	Detection Rate (%)	Number Detected	Detection Rate (%)
Newly diagnosed polyp/cancer	32	9	28	3	9
Inflammatory Bowel Disease	15	3	20	1	7
Healthy (anaemia or rectal bleeding)	22	11	50	3	14
Previous polyps/cancer	14	3	21	1	7
Total	83	25	30	8	9

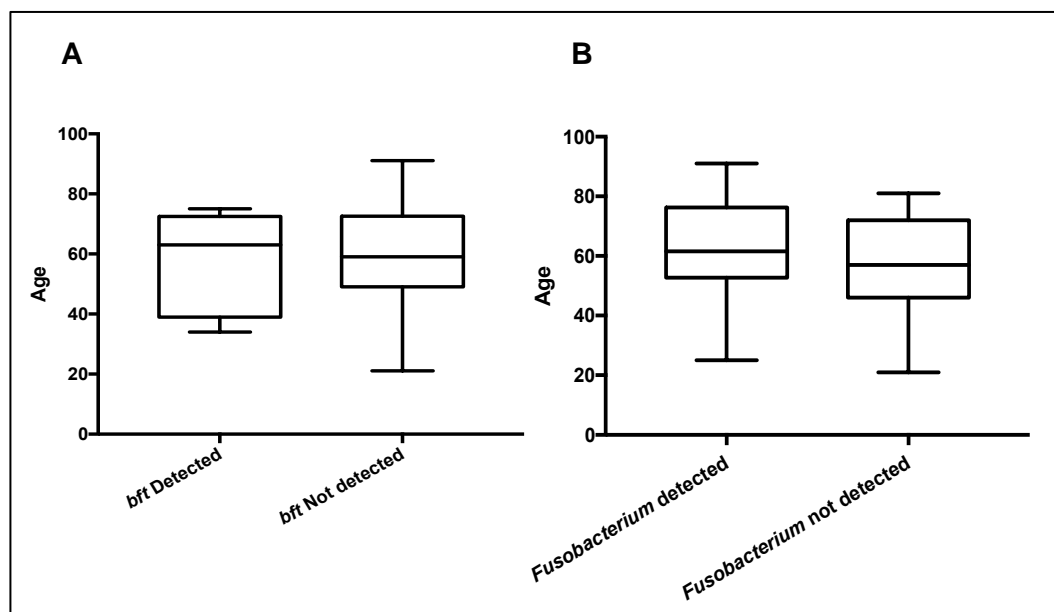
**4.2.5 Age and PCR positivity**

Due to the significant difference in mean age of individuals from different patient cohorts (Figure 4-1) and the significantly higher prevalence of *bft* in the newly diagnosed polyps/cancer groups compared with the healthy cohort (Figure 4-2) a comparison of the age of individuals positive and negative for PCR targets was carried out (Figure 4-4). Significant differences in age were not found between individuals positive or negative for *bft* (Figure 4-4 (A)) or the *Fusobacterium* genus (Figure 4-4 (B)).



**Figure 4-3 Percentage of individuals positive for *Fusobacterium*/*FadA***

PCR positivity was defined as one or more biopsy samples with a PCR amplicon for the specified gene target. Statistical significance was determined by Fisher's exact test.



**Figure 4-4 Comparison of age of individuals separated by *bft* (A) and *Fusobacterium* (B) positivity**

The box represents the interquartile range of the data with the horizontal line signifying the median. The range of the data is represented by the 'whiskers' of the box

(A) A box plot showing the age distribution in individuals PCR-positive or negative for *bft* patient groups.

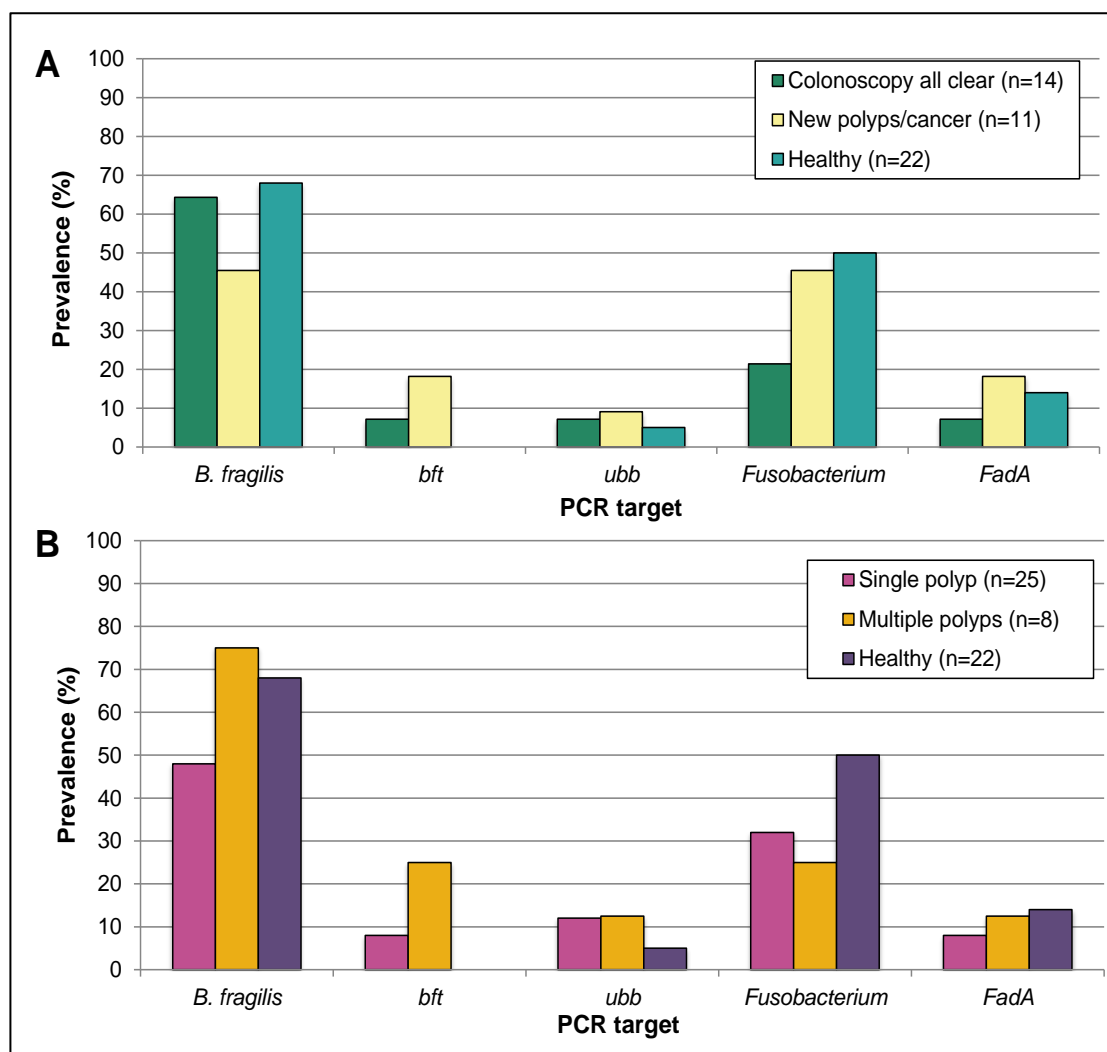
(B) A box plot showing the age distribution in individuals PCR-positive or negative for the *Fusobacterium* genus.

#### **4.2.6 Previous polyps/cancer**

The earlier defined cohort (section 2.1.1) referred to as having previous polyps/cancer had clear colonoscopies on their latest surveillance colonoscopy. Some individuals categorised in the newly diagnosed polyps/cancer group had previously had tumours removed. From all individuals with previous tumours a comparison was made between individuals with newly diagnosed polyps/cancer and individuals with an all clear colonoscopy (Figure 4-5A). *Bft* ( $p = 0.556$ ), *Fusobacterium* ( $p = 0.389$ ) and *fadA* ( $p = 0.566$ ) had a higher detection rate in the new polyps/cancer group but no significant differences were found.

#### **4.2.7 Individuals with multiple tumours**

Of potential interest when looking at a possible infectious link with cancer are individuals with more than one tumour. A comparison amongst individuals with a single tumour ( $n = 25$ ), multiple tumours ( $n = 8$ ) and the previously defined healthy group is shown in Figure 4-5B. The highest difference in detection rate between the single tumour and multiple tumour groups can be seen with *B. fragilis*, although this is not significant ( $p = 0.242$ ).



**Figure 4-5 Percentage of individuals positive for gene targets: Comparison of alternative patient groups**

(A) All individuals had previously had polyps/cancer. Prevalence is compared amongst individuals with a clear diagnostic colonoscopy, those with newly discovered tumours and the healthy cohort

(B) All individuals had newly diagnosed polyps. Prevalence is compared amongst individuals found to have a single tumour, multiple tumours and the healthy cohort.

#### 4.2.8 Cancer patients

In this study only three of the patients were diagnosed with a cancerous tumour as determined by histological analysis (Table 4-6). This study group is too small to make any statistical conclusion, but it should be noted that one individual was positive for *bft* and one positive for *ubb*, while two of the individuals were positive for *fadA*. *Bft* and *fadA*+ *Fusobacterium* was present in individual BF008, the only co-occurrence of these two gene targets found in this study.

**Table 4-6 PCR results of individuals with histologically defined cancer**

Individual	Cancer details	<i>B. fragilis</i>	<i>bft</i>	<i>ubb</i>	<i>Fusobacterium</i>	<i>FadA</i>
BF008	T3(b) sessile adenocarcinoma	+	+	-	+	+
BF035	Long-term IBD. High grade dysplasia adenocarcinoma	+	-	-	-	-
BF061	Rectosigmoid adenocarcinoma	+	-	+	+	+
Detection Rate (%)		100.0	33.3	33.3	66.7	66.7

**4.2.9 Culture-based identification utilizing selective media**

A plate wash PCR method using caecal effluent samples (section 2.6.3) was established to detect culturable bacterial isolates of interest. The two aims of this experiment were to investigate gene target prevalence with this different, luminal sample type and to generate stocks of isolates of interest for future analysis. To test for consistency of the method and batch to batch variation of the caecal effluent sample, repeat culturing and plate washing was carried out using five of the tested patient samples. All repeat samples gave the same PCR results as summarised in Table 4-6. A *bft* positive *B. fragilis* isolate was recovered from individual BF027 and from sequence analysis subtyped as *bft*-1.

Some differences in PCR results were seen between culturing from caecal effluent and DNA extracted from colonic biopsies (Table 4-7). The most discrepancies were seen with *Fusobacterium*, but was not detected in a preferred sample type.



**Table 4-7 Summary of plate wash PCR results from caecal effluent samples compared with biopsy PCR**

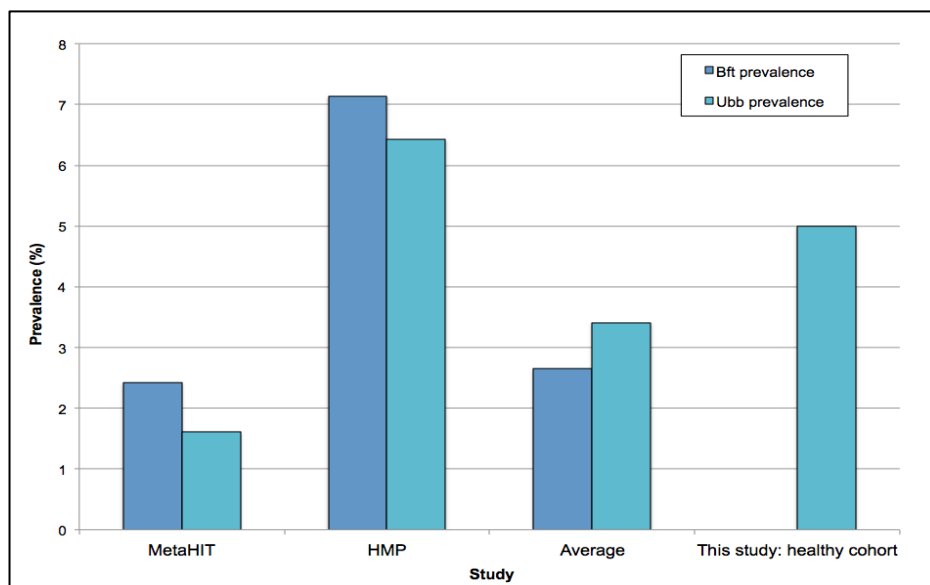
Patient ID	CE	Biopsy	CE	Biopsy	CE	Biopsy	CE	Biopsy	Patient group
	<i>B. fragilis</i>		<i>bft</i>		<i>Fusobacterium</i>		<i>fadA</i>		
BF027	+	+ (OT, NT, AT)	+	+ (OT, NT, AT)	+	-	N/A	-	Inflammatory polyp
BF028	+	+ (OT, NT, AT)	-	+ (OT, NT, AT)	+	-	+	-	Polyp
BF029	faint	+ (OT, NT, AT)	-	-	-	-	N/A	-	Polyp
BF030	-	-	-	-	+	-	N/A	-	Polyp
BF031	-	-	-	-	+	-	-	-	Healthy control
BF032	+	-	-	-	-	-	N/A	-	Polyp
BF033	+	+ (AC, DC)	-	+ (AC, DC)	-	-	N/A	-	IBD (UC)
BF035	-	-	-	-	-	-	N/A	-	IBD (UC)
BF036	-	-	N/A	-	-	+ (AC)	N/A	-	Healthy control
BF041	+	+ (AC, DC)	-	-	-	+ (AC)	N/A	-	IBD (Crohn's)
BF043	+	+ (AC)	+	+ (AC)	+	-	+	-	Polyp
BF046	faint	+	-	-	+	+ (AC, DC)	+	-	Healthy control
BF051	+	+	-	-	-	-	-	-	Healthy
BF056	+	-	-	-	-	-	N/A	-	IBD
BF057	+	+ (AC, DC)	-	-	-	-	N/A	-	Healthy
BF063	-	+ (AC, DC)	-	-	+	+ (AC)	+	+ (AC)	Healthy control
BF070	-	-	-	-	-	-	-	-	IBD
BF074	+	+ (AC, DC)	-	-	-	+ (AC, DC)	N/A	-	Healthy control
BF076	+	+ (AC, DC)	-	-	-	-	N/A	-	Previous polyp
BF086	+	+ (DC)	-	-	-	+	N/A	- (DC)	Healthy control
BF090	+	+ (AC, DC)	-	-	-	+ (AC, DC)	N/A	+ (AC, DC)	Polyp

Orange cells designate a difference between biopsy and CE PCR Result

CE = Caecal effluent, AC = Ascending colon, DC = Descending colon, OT = On tumour, NT = Next to tumour, AT = Away from tumour

### 4.2.10 Bioinformatic data mining

Data mining of two metagenomic studies, MetaHIT and the Human Microbiome Project (HMP), was carried out using KEGG BLAST (section 3.2.1) to compare prevalence results of *bft* and *ubb* from this study with other published results (Figure 4-6). The results demonstrate a similar *ubb* prevalence to the results presented. None of the individuals were positive for both *bft* and *ubb* genes.



**Figure 4-6 Prevalence of *bft* and *ubb* in metagenomic studies**

Sequences retrieved using a KEGG BLAST search. HMP = Human microbiome project

## 4.3 Discussion

### 4.3.1 *B. fragilis*

*B. fragilis* prevalence, even in healthy subjects, can vary widely between studies. Zitomersky *et al.*, 2013 found 73% of healthy individuals harboured *B. fragilis*, similarly to the results presented here, where 68% of the healthy cohort were PCR-positive for *B. fragilis*. While a study of the gut metagenomes, taken from stool samples, found 29% of subjects were *B. fragilis* positive in a metagenomic analysis (Chatzidaki-Livanis *et al.*, 2017)

### 4.3.2 *Bft*

In summary, *bft* was not detected in the healthy group, but *bft* prevalence was 7% in the currently healthy but previously had polyps/cancer group and 13% in both the IBD and newly

diagnosed polyps/cancer cohorts (Table 4-2). Previous research on *bft* epidemiology has varied in terms of detection method, sample type, sample location and patient cohorts (Table 4-8). Two previous studies found a significantly higher *bft* prevalence in CRC patients compared with healthy controls (38% vs 12% (Toprak *et al.*, 2006) and 27% vs 10% (Keenan *et al.*, 2016). In the same study Keenan *et al.*, 2016 found that culturing colonies on selective media versus DNA extraction gave differing results due to some samples having a low abundance of *bft* DNA. The limit of detection of the method may therefore be important when determining significant differences amongst cohorts but it is feasible that higher quantities of ETBF colonization is more clinically relevant. A higher *bft* prevalence of 40% than seen in other healthy cohorts was described by Zitomersky *et al.*, 2011.

The uppermost *bft* prevalence detected in a control group was 67% (Boleij *et al.*, 2015) although the control group contained many individuals with polyps found during colonoscopy. The same study also found cases (adenomatous polyps or CRC requiring surgical resection) to have a higher *bft* detection rate compared with controls. The prevalence of *bft* was also significantly associated with stage of CRC, a feature also seen in a study (Viljoen *et al.*, 2015) with a *bft* detection rate of 26% in CRC individuals. This may explain the lower prevalence seen in the results described here as the majority of tumours were adenomatous polyps, while out of the three individuals with cancer one was positive for *bft* (Table 4-6). The results described in this thesis found a significant difference in *bft* prevalence between the healthy/control and newly diagnosed colonic tumour group, in line with significant differences also found in the aforementioned studies.

**Table 4-8 Summary of the literature on *bft* epidemiology**

Study	Sample type	Method	Control group <i>bft</i> prevalence (%)	Polyps/CRC group <i>bft</i> prevalence (%)
(Toprak <i>et al.</i> , 2006)	Stool	Culture	12	38*
(Keenan <i>et al.</i> , 2016)	Stool	Culture, qPCR	10	27*
(Viljoen <i>et al.</i> , 2015)	Biopsies	DNA extraction and qPCR	N/A	26
(Boleij <i>et al.</i> , 2015)	Biopsies	Culture and touch-down PCR	67	89
(Purcell <i>et al.</i> , 2017)	Biopsies	DNA extraction and qPCR	N/A	50

\* statistically significant

### 4.3.3 *Ubb*

This study found no significant differences in detection rate amongst patient groups, but the newly diagnosed polyps and IBD group had higher detection rates, perhaps meriting a larger study. The only comparative study for *ubb* prevalence mined the integrated gene catalogue of gut (stool) metagenomes and found *ubb* to be present in 3.47% of metagenomes, a number close to the 5% prevalence in the healthy cohort described here. The two IBD individuals positive for *ubb* were also positive for *bft*. Whole genome sequencing of the BOB25 (Nikitina *et al.*, 2015) clinical isolate showed that this strain contains both *ubb* and *bft*, it is therefore possible that the individuals with IBD harbour a similar strain. A recently published study (Chatzidaki-Livanis *et al.*, 2017) has identified BfUbb as inhibiting the growth of other *B. fragilis*, which may promote chronic colonisation of the strain.

### 4.3.4 *Fusobacterium*, *fadA* and CRC

A consistent finding in studies of the CRC microbiome is the enrichment of *Fusobacterium* species on tumour tissue in comparison to adjacent healthy mucosa (Marchesi *et al.*, 2011, Kostic *et al.*, 2012, Castellarin *et al.*, 2012, Tahara *et al.*, 2015). Whilst no significant differences in regard to *Fusobacterium* and *fadA* were observed in this study, the results here were measuring prevalence in individuals newly diagnosed with tumours compared with a healthy cohort, as opposed to abundance within an individual. The majority of tumours collected in this study were adenomatous polyps as opposed to cancerous tissue. In fact, of the three individuals with CRC two were positive for *Fusobacterium* and *fadA*.

#### Chapter 4.3.5 *Fusobacterium* and IBD

Unexpectedly, the healthy cohort had a high prevalence of *Fusobacterium*. One theory to explain this finding is that the *Fusobacterium* primers have not been tested against *Cetobacterium somerae*, which has a closely related 16S rRNA gene sequence and has been shown to be present in healthy biopsy tissue (Drewes *et al.*, 2017). Another explanation is that there are gastrointestinal commensal *Fusobacterium* species and distinctive species that are associated with CRC.

*Leptotrichia* is a genus closely related to the genus *Fusobacterium* and together are in the class *Fusobacteria*. *Leptotrichia* species can also make the adhesin FadA, considered a virulence factor in *Fusobacterium*-associated carcinogenesis (Rubinstein *et al.*, 2013). The FadA primers used here are specific for *F. nucleatum* and *F. periodonticum*, possibly primers designed to capture *F. nucleatum* and *Leptotrichia* FadA would give a different outcome, particularly as *Leptotrichia* species have been found associated with CRC tumour tissue (Warren *et al.*, 2013).

#### **4.3.5 *Fusobacterium* and IBD**

No significant difference in the detection rate of either *Fusobacterium* or FadA was found between the healthy cohort and individuals with IBD. This is in contrast to some of the published literature. Increased levels of the genus *Fusobacterium* have been found in new onset paediatric Crohn's (Gevers *et al.*, 2014), Crohn's (Pascal *et al.*, 2017) and IBD patients with PSC (Sabino *et al.*, 2016). IBD is a heterogeneous disease, differing in the extent of radial inflammation, location and age of onset. The aforementioned studies focused on a specific type of IBD whereas the IBD cohort in this thesis comprises individuals with UC and Crohn's, and with different levels of severity.

*Fusobacterium* positive individuals from the IBD group had either previous polyps or new polyps. Long-term inflammation in IBD patients is linked to an elevated risk of developing CRC (Kim and Chang, 2014). The degree of inflammation correlates to colorectal neoplasia in UC (Rutter *et al.*, 2004) but it is possible that a bacterial pathogen could be driving this inflammation. Colitis associated cancer has been shown to have a different microbiome to sporadic CRC but interestingly *Fusobacterium* is one of these differentiating microbiomes associated with sporadic but not colitis associated cancer (Richard *et al.*, 2017). *F. varium* had been identified as a putative UC pathogen (Ohkusa *et al.*, 2003) capable of invading cells and

#### Chapter 4.3.6 *B. fragilis*, *bft* and IBD

initiating an inflammatory response. *F. nucleatum* is also intracellular with strains isolated from inflamed mucosa shown to be more invasive than those isolated from non-affected tissue (Strauss *et al.*, 2011). *Fusobacterium* association with IBD may also be connected with geographical location as a different study found *F. mortiferum* to be the most common *Fusobacterium* species isolated from an IBD cohort (Lee *et al.*, 2016b).

Literature on the microbiome in colitis associated polyps and cancer is limited. Microbes could be key drivers of tumorigenesis in IBD patients. In addition, there is scope for using the microbiome as a biomarker to identify IBD patients at risk of developing CRC.

#### **4.3.6 *B. fragilis*, *bft* and IBD**

Literature on the association of *bft* and IBD has produced mixed results. The results presented in this thesis found a higher prevalence of *bft* in the IBD cohort although these results were not significant. A statistically significant difference in *bft* prevalence was found using a culture-based method from UC biopsy samples (Zamani *et al.*, 2017) by one study but another study found no significant difference (Basset *et al.*, 2004). Similar prevalence values to my study were found by Prindiville *et al.*, 13.5% in the IBD group versus 2.89% in the control group (Prindiville *et al.*, 2000). However, they studied stool samples in comparison to the more biologically relevant mucosal biopsy samples used in this thesis. Another difference is that the IBD cohort included in this thesis were recruited if there were no signs of macroscopic inflammation during colonoscopy, whereas Prindiville *et al.*, and Basset *et al.*, found associations between *bft* and active disease. A possible key factor is that Zamani *et al.*, specifically looked at UC, whereas Basset and this thesis investigated a variety of IBD types, with and without active disease.

Looking at the species level there is an agreement between Zitomersky *et al.*, Zamani *et al.*, and the results presented here in that there is no significant difference in *B. fragilis* prevalence between control and IBD groups. This finding and the mixed *bft* results make an argument against using 16S rRNA sequencing to investigate microbial dysbiosis in IBD.

Two individuals positive for *bft* (section 4.2.3) had inflammatory pseudopolyps but did not have IBD. Inflammatory polyps are a rare finding for individuals without colitis, but provides an anecdotal link between inflammation, tumorigenesis and *bft*.

## Chapter 5 : ANALYSIS OF BACTEROIDES FRAGILIS POLYSACCHARIDE A IN VIVO

### 5.1 Introduction

#### 5.1.1 Antigenic and phase variation

Phenotypic variation within a bacterial population can be manifested by phase variation, a method used by bacteria to switch expression of an antigen on and off. Often one antigen will be switched off by phase variation and an alternative version is switched on to create antigenic variation. This phenomenon of phase variation was first described in 1922 in *Salmonella typhimurium* flagellar antigens by Andrewes (Andrewes, 1922). The classic theory for antigenic variation is as a strategy to avoid specific immune responses. Various genetic and epigenetic regulatory mechanisms exist to create antigenic variation, described in detail in Van der Woode *et al.*, 2004 with common mechanisms summarised in Table 5-1.

**5-1 Common mechanisms of antigenic variation**

Mechanism	Description	Example
Homologous recombination	Multiple copies of incomplete antigen gene variants except the expressed variant, for example no transcriptional promoters. DNA recombination results in functional, chimeric gene formation.	<i>Neisseria gonorrhoeae</i> pilin genes (Segal <i>et al.</i> , 1986)
Slipped-strand mispairing	Misalignment of repeat DNA sequence during replication. Can be found in promoter regions with the consequence of altered efficiency of transcription or if in coding sequence causing an altered reading frame.	TA repeats in <i>Haemophilus influenzae</i> fimbriae promoters encoded by <i>hif</i> genes (van Ham <i>et al.</i> , 1993)
DNA inversion	Inversion of a DNA element, flanked by inverted repeat sequences, by site-specific recombination.	<i>B. fragilis</i> capsular polysaccharides (see section 5.1.2)

*Borrelia burgdorferi* is the causative agent of Lyme disease in humans, spread by a tick vector. These spirochetes can infect humans for a long period of time and as result have a complex mechanism of antigenic variation as a method of immune evasion. The bacterium has a highly immunogenic surface antigen (Lawrenz *et al.*, 2004), Vmp-like sequence (VlsE) which is hypothesised to stimulate a strong antibody response in order to detract from other surface

antigens. Throughout infection novel variants of VlsE are generated via segmental recombination between the *VlsE* expression site and many *vls* silent gene cassettes (Zhang *et al.*, 1997).

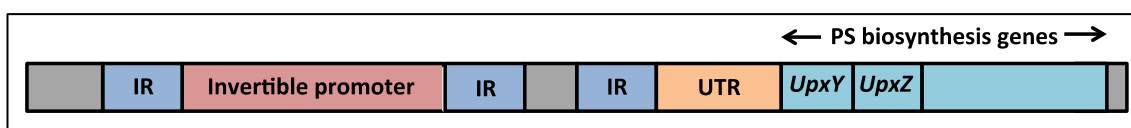
### **5.1.2 Capsular Polysaccharide (PS) biosynthesis and transcription**

Bacterial capsules are a layer, frequently comprised of polysaccharide, outside the cell envelope. Capsules are used by bacteria as a strategy to evade the host immune system, for example by preventing the formation of a complement attack complex, and to prevent bacteriophage attack. Capsules are also used for adhesion and colonisation and for these reasons are considered virulence factors

*B. fragilis* has three distinct capsule layers. Each strain has either a small or large antigenically different fibrous outer layer and a narrow, electron dense layer also known as the microcapsule (Kasper *et al.*, 1977). This variable surface microcapsule (MC) is located outside the cell envelope, is co-expressed with the large capsule and is comprised of different capsular polysaccharides (PS). DNA sequencing of the *B. fragilis* type strain revealed eight different capsular polysaccharide biosynthetic loci designated PSA-H (Krinos *et al.*, 2001, Cerdeno-Tarraga *et al.*, 2005). Genomic comparison of further *B. fragilis* strains showed an unpredicted diversity in PS biosynthetic loci (Patrick *et al.*, 2010) with a total of 28 types discovered to date. Each type of capsular PS has its own locus of PS biosynthesis genes for an individual PS type (Comstock *et al.*, 1999). Upstream of each of the polysaccharide biosynthetic loci is a single promoter region. With the exception of PSC the promoters can undergo inversion of their DNA (Krinos *et al.*, 2001). In fact a characteristic of the *B. fragilis* genome observed as a result of whole genome sequencing is the large number and variety of DNA inversions (*fragilis* invertible regions, *fin*) present in the genome (Kuwahara *et al.*, 2004, Cerdeno-Tarraga *et al.*, 2005). The bacterium is capable of switching on and off PS synthesis to alter capsular PS expression in a phenotypic switching process known as phase variation. The mechanism of *B. fragilis* surface PS phase variation is by DNA inversion of the PS biosynthesis promoter (Krinos *et al.*, 2001). In seven of the eight PS biosynthetic loci the region of invertible DNA is flanked by repeat sequences known as *fragilis* inversion crossover sites (*fix*) (Patrick *et al.*, 2003). Promoter DNA inversion is mediated by a serine site-specific recombinase designated as multiple promoter invertase, Mpi (Coyne *et al.*, 2003).



The first gene of all eight PS biosynthetic loci is the *upxY* gene where x is replaced by a to h depending on the specific PS locus (Figure 5-1). The UpxY family of proteins share homology amongst individual proteins, but contain two regions of location specific (a-h) amino acid sequence (Chatzidaki-Livanis *et al.*, 2009). The UpxY proteins are able to associate with RNA polymerase in the 5' untranslated region (UTR) to prevent premature termination of transcription. The adjacent *upxZ* genes code for a family of proteins able to prevent the transcriptional antitermination function of other PS loci UpxY proteins. Altogether the UpxZ proteins prevent simultaneous synthesis of PS types in a single bacterial cell by a hierarchical system of regulation with PSC the default locked 'ON' promoter (Chatzidaki-Livanis *et al.*, 2010).



**Figure 5-1 Schematic of the PS promoter and biosynthetic locus**

IR = Inverted repeat, UTR = Untranslated region, PS = Polysaccharide Based on (Coyne *et al.*, 2001)

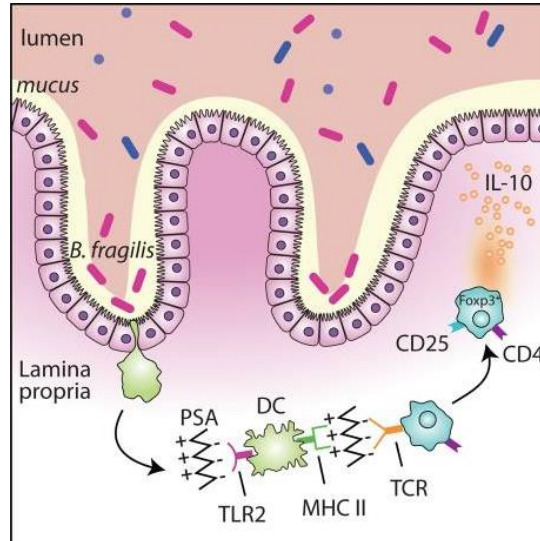
### 5.1.3 Polysaccharide A (PSA)

The most studied of the capsular polysaccharides is PSA, which is arguably the most immunogenic and important. The structure of PSA comprises a tetrasaccharide repeating unit, which unusually for a capsular polysaccharide is zwitterionic, comprising a positive amino group and a negative carboxyl group (Baumann *et al.*, 1992). The repeating unit makes up a right-handed helix with the charged groups externally facing allowing for interaction with other molecules (Wang *et al.*, 2000). PSA has been studied extensively due to its immunomodulatory properties, in part due to the zwitterionic nature and structure of this molecule.

Figure 5-2 illustrates the mechanism of PSA immune stimulation which originates with the interaction of PSA with TLR2 on dendritic cells (DCs) (Round *et al.*, 2011, Dasgupta *et al.*, 2014). DCs process PSA where it is presented on Major Histocompatibility Complex Class II (MHC II) (Duan *et al.*, 2008) molecules that interact with T-cell receptors (TCRs). The consequence of this interaction is an expansion of interleukin 10 (IL-10) producing CD4<sup>+</sup> cells (Mazmanian *et al.*, 2005). IL-10 is an anti-inflammatory cytokine and an increase in its

#### Chapter 5.1.4 Restriction-modification systems

production by PSA can mediate the prevention of *H. hepaticus*-induced colitis in mice (Mazmanian *et al.*, 2008). PSA signalling via TLR2 also suppresses Th17 and interleukin-17 (IL-17) production to promote mucosal colonisation of *B. fragilis* (Round *et al.*, 2011).



**Figure 5-2 Mechanism of immune cell activation**

Zwitterionic *B. fragilis* PSA can activate toll-like receptor 2 (TLR2) signalling in dendritic cells. PSA is presented on MHC class II molecules where it can interact with T-cell receptors (TCR) to trigger CD4 Foxp3<sup>+</sup> T cell mediated IL-10 production. Figure taken from (Surana and Kasper, 2012).

#### 5.1.4 Restriction-modification systems

These systems act as bacterial innate immune systems to guard against invading bacteriophage. They comprise a restriction endonuclease (REase) that recognises and cleaves specific foreign DNA sequences and a methyltransferase (MTase) to discriminate between self and non-self DNA (Vasu and Nagaraja, 2013). The property of restriction enzymes to recognise specific DNA sequences has been utilised in many molecular biology techniques such as molecular cloning and restriction fragment length polymorphism (RFLP), used in DNA fingerprinting and profiling. This chapter uses a previously described method (Krinos *et al.*, 2001, Troy *et al.*, 2010) using a restriction enzyme to asymmetrically digest DNA within the invertible PSA promoter thereby generating different sized DNA fragments depending on the orientation of the promoter.

The work in this chapter describes the first analysis of PSA promoter orientation *in vivo* and analyses PSA promoter orientation in *B. fragilis* populations from different patient cohorts. Possible downstream implications of PSA production/absence were subsequently investigated.

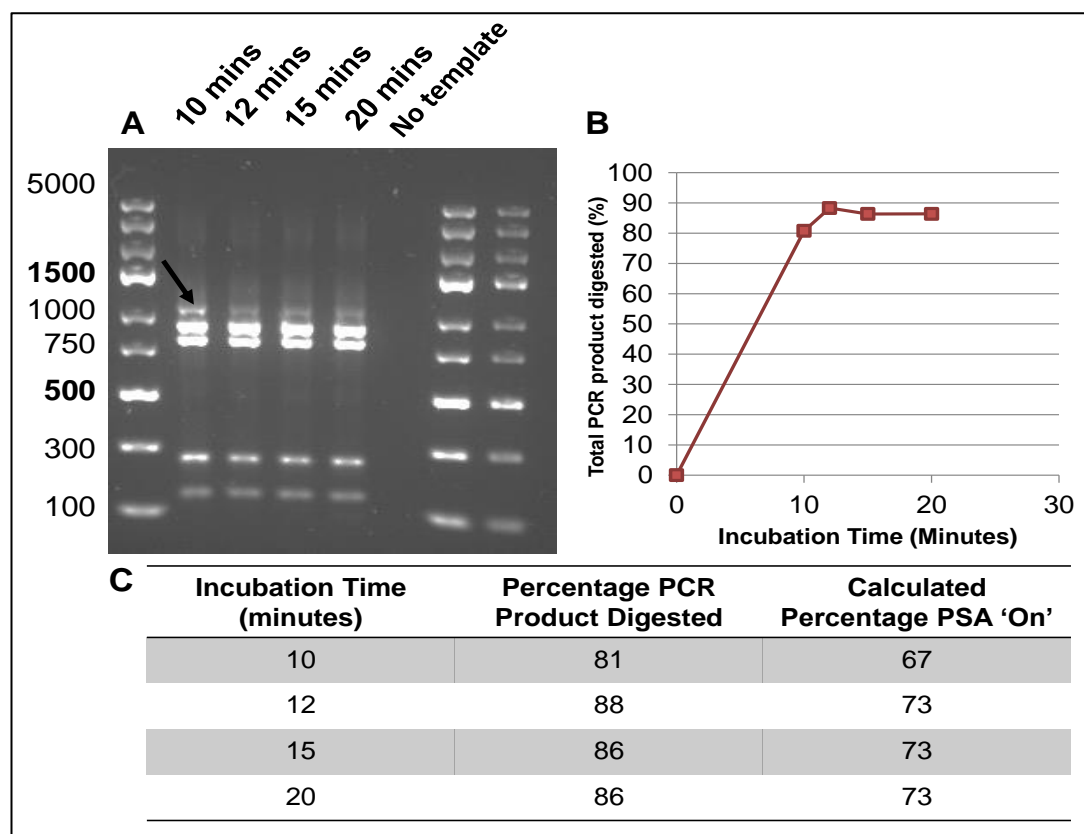
## 5.2 Results

### 5.3 Polysaccharide A (PSA) promoter orientation

#### 5.3.1 Restriction Digest optimization

A method to determine PSA promoter orientation has previously been published (Krinos *et al.*, 2001, Troy *et al.*, 2010). This method was replicated in the laboratory using a 'colony' PCR method on a *B. fragilis* type strain NCTC 9343 colony. However, it was found that the manufacturer's recommended incubation time for the SspI restriction enzyme (10 min) led to incomplete digestion of the PCR product.

A series of restriction enzyme incubation times were tested and the extent of PCR product digestion determined by resolving on an agarose gel and measuring band intensity (Figure 5-3). An incubation time longer than 10 minutes increased the total amplicon digestion to a plateau of ~86-88%, so an incubation time of 20 min was chosen for subsequent experiments.



**Figure 5-3 Optimisation of Restriction Digest Time for the PSA promoter orientation assay**

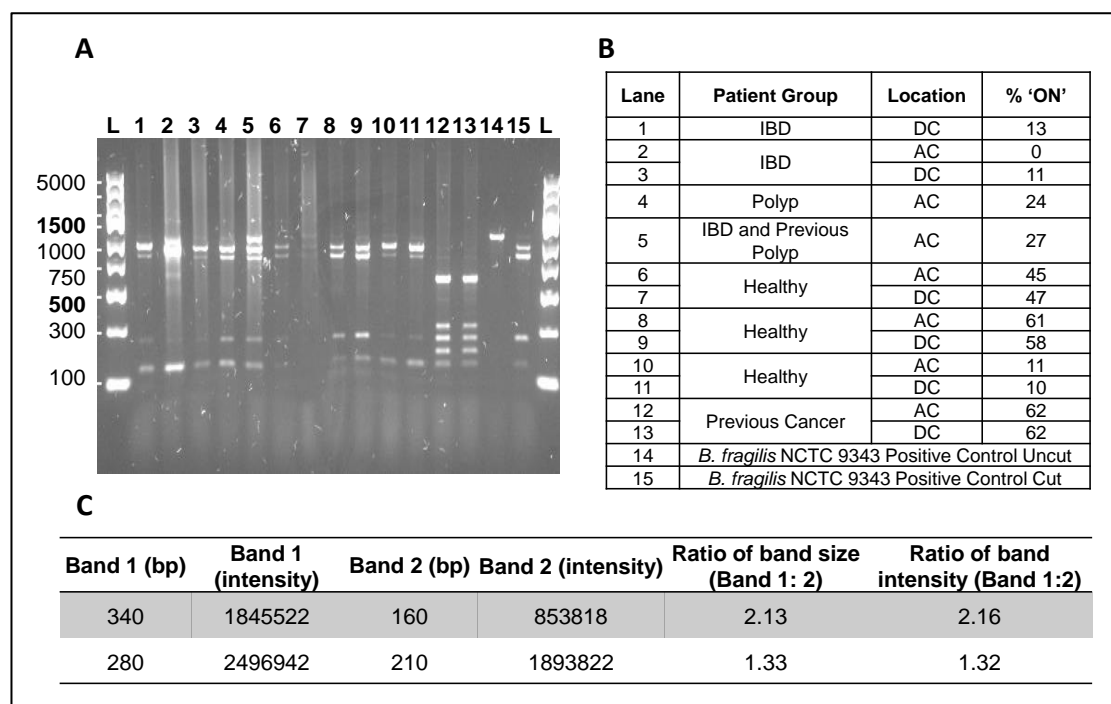
(A) Agarose gel depicting PCR products digested with SspI at different incubation times. Each lane represents a different incubation time. The bold arrow indicates the undigested PCR product. GeneRuler Express ladder marker sizes are given to the left of the gel with brighter reference bands labelled in bold.

(B) Comparison of the total PCR product digestion with restriction enzyme incubation time

(C) Table to compare the percentage of PCR product digested with the calculated value for PSA promoter 'ON'

### 5.3.2 Identification of a Single Nucleotide Polymorphism (SNP) in UpaY

Following optimisation of the method, extracted colonic biopsy DNA (Section 2.2.1) was used to determine the orientation percentage of the *B. fragilis* PSA promoter in patient samples. Whilst carrying out the PSA restriction digest protocol on clinical samples, it was observed that some samples had a different SspI digest pattern. Lanes 12 and 13 in Figure 5-4A show samples with two additional bands at approximately 200 and 250 bp. Published results using this method have previously used DNA isolated from the *B. fragilis* type strain (Krinos *et al.*, 2001, Troy *et al.*, 2010), and so this digest pattern had not previously been observed.



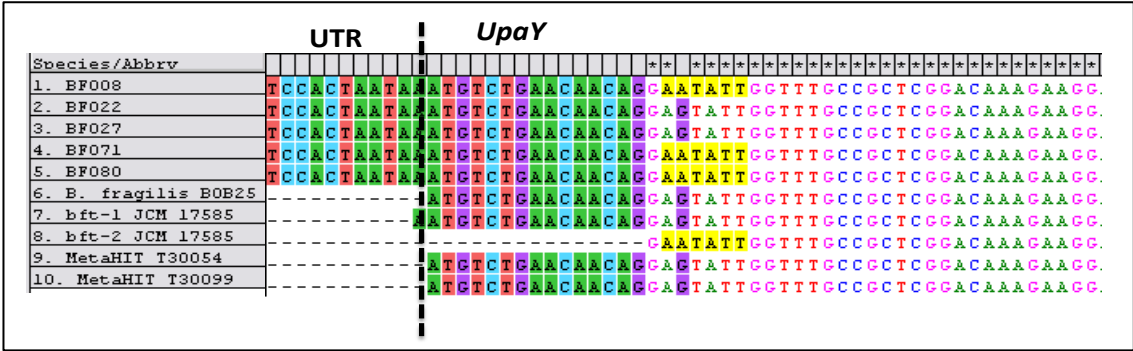
**Figure 5-4 Example Restriction Digest DNA Gel Electrophoresis**

(A) PSA PCR products from extracted biopsy DNA digested with SspI restriction enzyme. Individual samples (Figure 1-2B) were loaded per lane. GeneRuler Express ladder marker sizes are given to the left of the gel with brighter reference bands labelled in bold.

(B) Characteristics of the sample and calculated PSA promoter % orientated 'ON'. Colon location of the biopsy sample are listed as AC = ascending colon and DC = descending colon.

(C) Table outlining the band characteristics in lane 12 of the unusual digest pattern. Calculations were used to determine the promoter orientation of the individual bands.

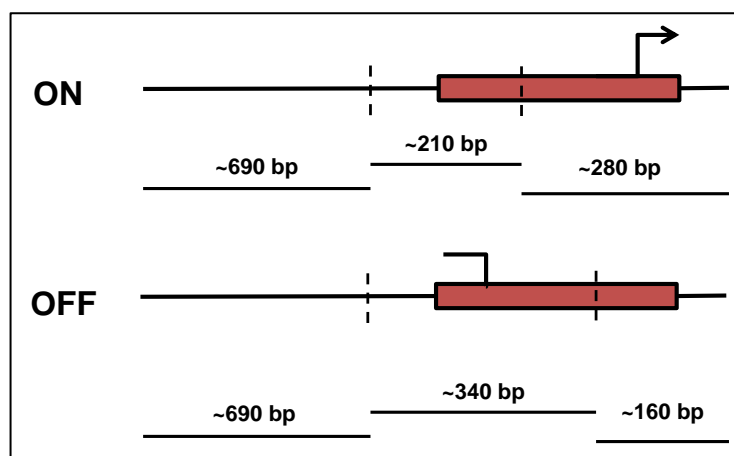
The undigested PCR products from select samples with and without the alternative digest pattern were purified and sent for DNA sequencing. DNA sequences (5' to 3') were aligned against published sequences from other *B. fragilis* isolates and sequences from the MetaHIT (Li *et al.*, 2014) metagenomic project (Figure 5-5). The alternative digest pattern was found to be due to an additional SspI recognition sequence in some of the isolates due to a single nucleotide polymorphism (SNP). The codon changes from GAG to GAA, although this will still code for a glutamic acid amino acid residue. The SNP is located close to the start of the first gene (*upaY*) of the PSA biosynthetic loci as highlighted in Figure 5-5.



**Figure 5-5 Polysaccharide A (PSA) Promoter Region DNA Sequences from Different *B. fragilis* strains**

DNA sequences aligned in MEGA of PSA Promoter PCR products from patient samples, published sequences and sequences found from the MetaHIT metagenome project. The additional SspI restriction enzyme cut sequence in some isolates is highlighted in yellow. UTR = Untranslated region

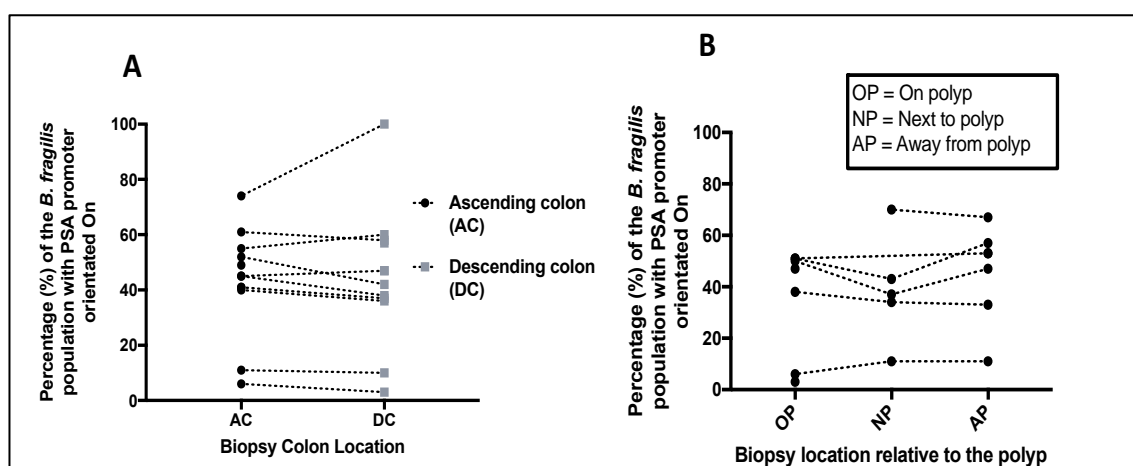
Using a combination of DNA gel electrophoresis digest pattern (Figure 5-4A) and DNA sequences (Figure 5-5) the size and orientation of restriction digest fragments in isolates with the extra SspI restriction site was determined and is illustrated in Figure 5-6. Following separation of restriction digest fragments on an agarose gel the DNA intercalating dye will decrease in intensity as the fragment size decreases. This concept was used to determine the promoter orientation of the alternatively cut bands (Figure 5-4C), using lane 12 in Figure 5-4A as an example. The percentage of the *B. fragilis* population promoter orientation could then be calculated. The method was applied to biopsy DNA isolated from clinical samples that had previously be found to harbour *B. fragilis* (Figure 5-2A). The SNP was found in individuals from all patient groups (1 x healthy control, 2 x IBD, 1 x newly diagnosed polyp, 1 x newly diagnosed cancer, 1 x previous polyps and 1 x previous cancer).



**Figure 5-6 Schematic of Restriction digest fragments of the Polysaccharide A (PSA) promoter regions with extra SspI restriction site**

The invertible promoter region of the PCR product is coloured in red. SspI cut sites are marked by a dashed line with the additional cut site outside the invertible region.

### 5.3.3 PSA promoter orientation separated by biopsy location



**Figure 5-7 Percentage of *B. fragilis* cells with the Polysaccharide A (PSA) promoter in the 'On' orientation: comparison between colonic biopsy samples**

(A) Comparison of percentage PSA promoter orientated on in biopsy sample *B. fragilis* populations obtained from different colon locations within the same individual. AC = ascending colon. DC = descending colon. Samples from the same individual are paired with a straight line. A Wilcoxon matched-pairs signed rank test was used to statistically compare biopsy locations.

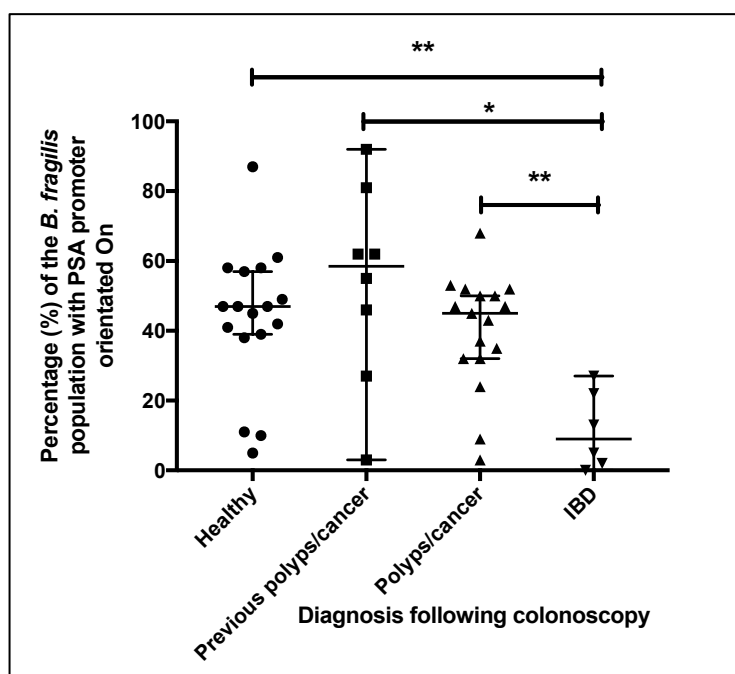
(B) Comparison of percentage PSA promoter orientated on in biopsy sample *B. fragilis* populations obtained from different colon locations relative to the tumour within the same individual. OP = On the polyp colon. NP = Next to the polyp (2 cm away) and AP = Away from the polyp (10 cm away). Samples from the same individual are paired with a straight line. A Wilcoxon matched-pairs signed rank test was used to statistically compare biopsy locations.

The percentage of the *B. fragilis* population with the PSA promoter orientated on was not significantly different in samples from the same individual taken from the ascending and

descending colon (Figure 5-7A) as determined by a paired, two-tailed t test ( $p = 0.976$ ). An average percentage of all samples obtained from the same individual were used in subsequent analyses.

### 5.3.4 PSA promoter orientation separated by patient groups

Individuals were divided into groups according to the results of the diagnostic colonoscopy. Figure 5-8 shows that individuals with IBD had a significantly lower proportion of *B. fragilis* with PSA orientated on compared with healthy individuals, those previously with colonic polyps/cancer and those with newly diagnosed polyps/cancer. F-tests between patient groups determined that there was no significant difference between the variances of the different patient group populations.



**Figure 5-8 Percentage of *B. fragilis* cells with the Polysaccharide A (PSA) promoter in the 'On' orientation: comparison amongst patient groups**

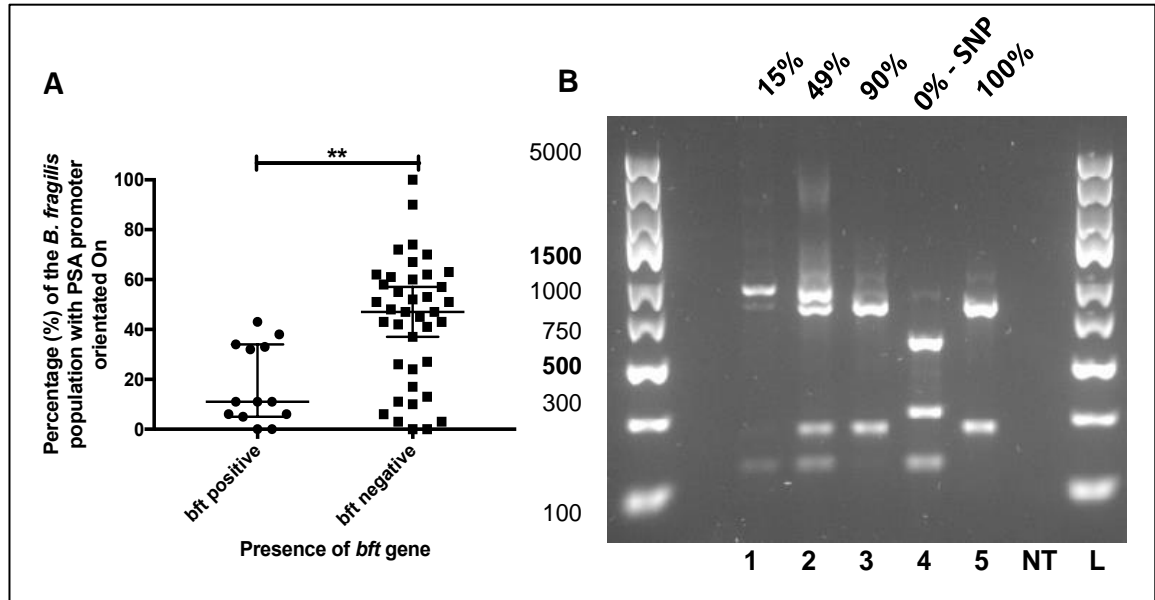
Significant differences amongst patient groups was determined using the non-parametric Kruskal-Wallis test and pairwise difference between patient groups established using Dunn's post-test multiple comparisons correction (Table 5-3). Error bars indicate the median and 95% confidence level. A \* designates a  $p$  value of less than 0.05 and \*\* represents a  $p$  value of less than 0.01.

Figure 5-9(A) shows that that *bft* positivity was significantly associated ( $p < 0.000$ ) with a low percentage of the *B. fragilis* population with the PSA orientated on. To determine whether this was a feature of *bft* positive *B. fragilis* isolates, the percentage of the *B. fragilis* population with



#### Chapter 5.3.4 PSA promoter orientation separated by patient groups

the promoter orientated on was determined for a selection of *B. fragilis* isolates grown *in vitro* on agar (Figure 5-9B). The PSA promoter orientation on the other hand had no significant difference ( $p = 0.826$ ) with the presence of *ubb*.



**Figure 5-9** Individuals with *bft* positive strains have a significantly lower percentage of *B. fragilis* with the PSA promoter in the 'ON' position

(A) Percentage of *B. fragilis* cells with the PSA promoter in the 'ON' orientation as separated by individuals with biopsy samples positive for the *B. fragilis* toxin gene (*bft*). Mann-Whitney U test was used determined significant differences between *bft* positive and negative samples. Error bars indicate the median and 95% confidence level .

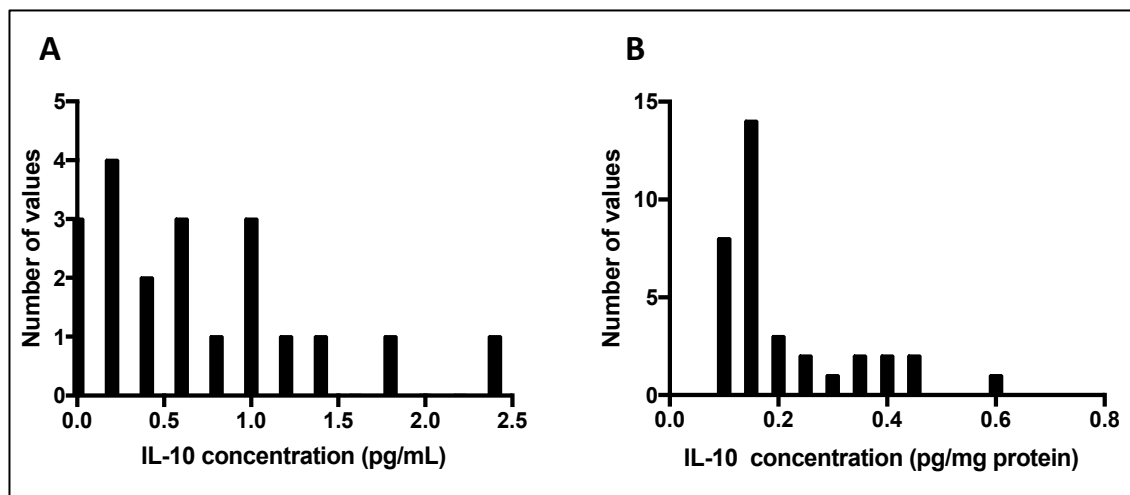
(B) Each lane represents a different *B. fragilis* isolate (1) = BF001.DC, (2) = NCTC 9343, (3) = JCM 17585 bft-1, (4) = JCM 17586 bft-2, (5) = JCM 17587 bft-3, (NT) = No template control, (L) = GeneRuler Express DNA ladder. DC = Descending colon. The calculated percentage of *B. fragilis* with the PSA promoter orientated 'ON' is provided for each isolate at the top of the relevant lane.

## 5.4 Interleukin 10 (IL-10)

Following the significant PSA promoter orientation differences amongst patient groups, it was decided to investigate the relationship with IL-10 levels as the most probable downstream consequence of immunomodulatory PSA (Figure 5-2). Non-parametric statistical tests were employed, as the results were not normally distributed as shown in Figure 5-10.

### 5.4.1 Serum

The frequency distribution of IL-10 concentration in serum samples from different individuals showed a subset of samples having higher IL-10 concentrations, while many of the samples had values that were undetectable (Figure 5-10A).

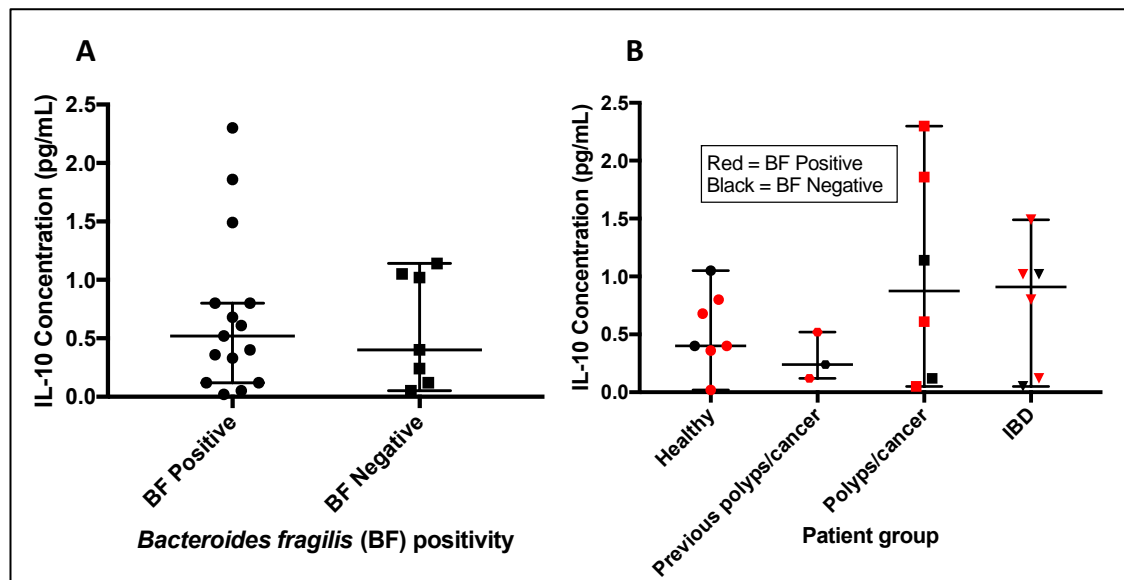


**Figure 5-10** Frequency distributions of serum and tissue IL-10 concentration do not follow a normal distribution

(A) Histogram to show the distribution of serum IL-10 concentrations (pg/mL)

(B) Histogram to show the distribution of tissue IL-10 concentrations (pg/mg protein)

No significant difference in serum IL-10 concentration ( $p = 0.932$ ) was observed between individuals with positive or negative PCR results for *B. fragilis* (Figure 5-11A). In addition, no significant difference in serum IL-10 concentration was detected amongst patient groups (Figure 5-11B). The highest serum IL-10 concentrations were obtained from individuals (BF001, BF005, BF090) with positive *B. fragilis* results and polyps/IBD. These individuals had long-term, but currently quiescent UC; polyp (unknown type); and a tubular adenoma with diverticular disease.



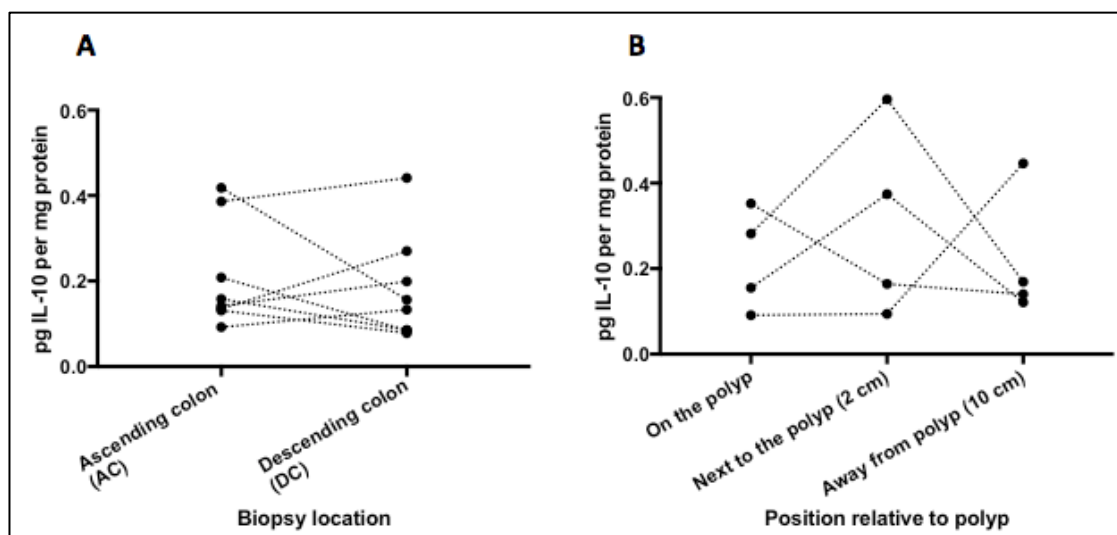
**Figure 5-11 Serum IL-10 concentration separated by patient groups or between *B. fragilis*-positivity**

(A) Levels of serum IL-10 concentration separated by individuals with biopsy samples PCR-positive or negative for *B. fragilis*. Statistical significance was calculated using a Mann-Whitney U test.

(B) Levels of serum IL-10 concentration separated by patient group. Significant differences amongst patient groups were determined using the non-parametric Kruskal-Wallis test and pairwise differences between patient groups calculated using Dunn's post-tests multiple comparisons correction (Table 5-3). The colour red indicates individuals with *B. fragilis* positive samples. Error bars for figures A and B indicate the median IL-10 concentration (pg/mL) and 95% confidence level (C.I.).

#### 5.4.2 Tissue lysate

As with PSA promoter orientation (Figure 5-12), there was no significant difference in IL-10 concentration amongst colonic biopsies collected from different locations within the same individual. IL-10 concentrations in locations relative to the polyp show slight variation within an individual but this was not significant, and no consistent trend could be observed.

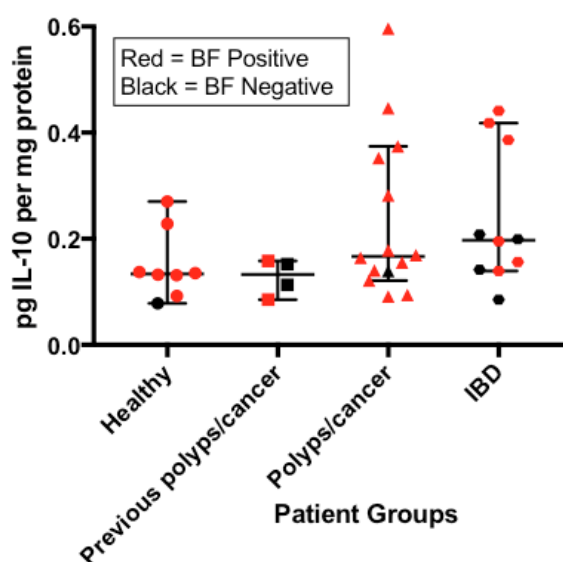


**Figure 5-12 Tissue IL-10 concentration amongst biopsy sites**

Dotted lines link samples from the same individual

(A) Levels of tissue IL-10 concentration in the ascending colon (AC) and descending colon (DC) from the same individual. A Wilcoxon matched-pairs signed rank test was used to statistically compare IL-10 levels from different biopsy locations.

(B) Levels of tissue IL-10 concentration at sites relative to a polyp/cancer. Statistical analysis amongst biopsy sites was determined using the Wilcoxon matched-pairs signed rank test.



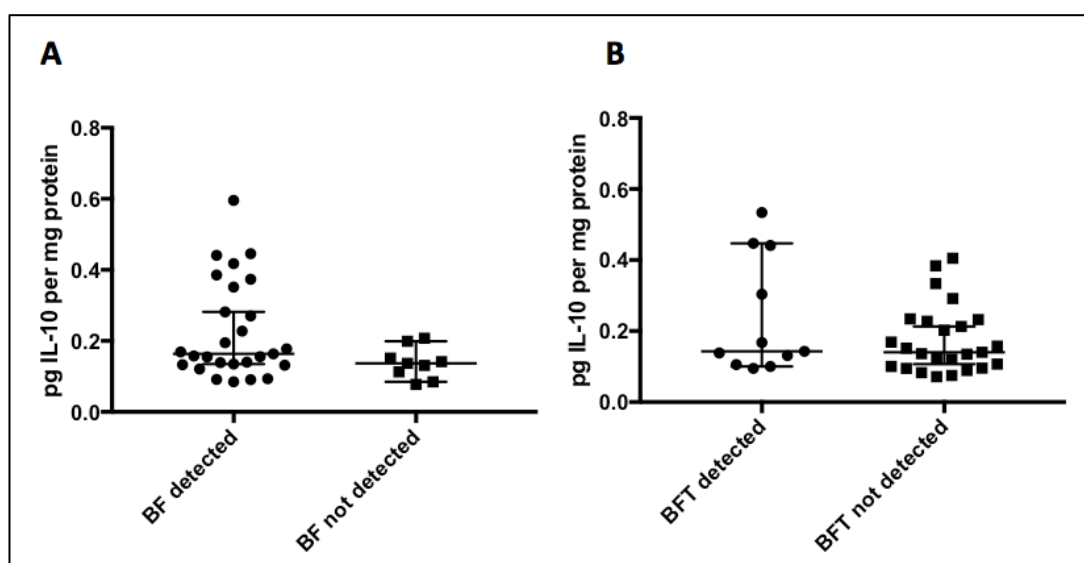
**Figure 5-13 Tissue IL-10 concentration amongst patient groups**

Differences in IL-10 levels in the different patient groups were determined using the non-parametric Kruskal-Wallis test and pairwise difference between patient groups established using Dunn's post-test multiple comparisons correction (Table 5-3). The colour red indicates individuals with *B. fragilis* positive samples. Error bars indicate the median and 95% confidence level.

#### Chapter 5.4.2 Tissue lysate

When individuals were separated according to patient group there was no significant differences in mean pg IL-10 per mg protein (Kruskal-Wallis  $p = 0.123$ ) amongst patient groups (Figure 5-13). Similarly, the Dunn's multiple comparisons test did not identify (Table 5-2) any significant differences in IL-10 quantities between pairs of patient groups. The three individuals with the highest tissue IL-10 concentrations were obtained from individuals, BF006, BF008 and BF001, with sigmoid polyp, ascending colon adenocarcinoma and quiescent UC respectively (Figure 5-11).

The median and 95% C.I. of tissue IL-10 concentration (Figure 5-14A) was higher in *B. fragilis* positive samples (0.164) compared to negative samples (0.137) although this difference was not significant ( $p = 0.069$ ). However, BFT positive biopsy samples (Figure 5-14B) had a very similar median (0.143) compared with BFT negative samples (0.140), a difference with a larger  $p$  value as determined by the Mann-Whitney U test ( $p = 0.3548$ )



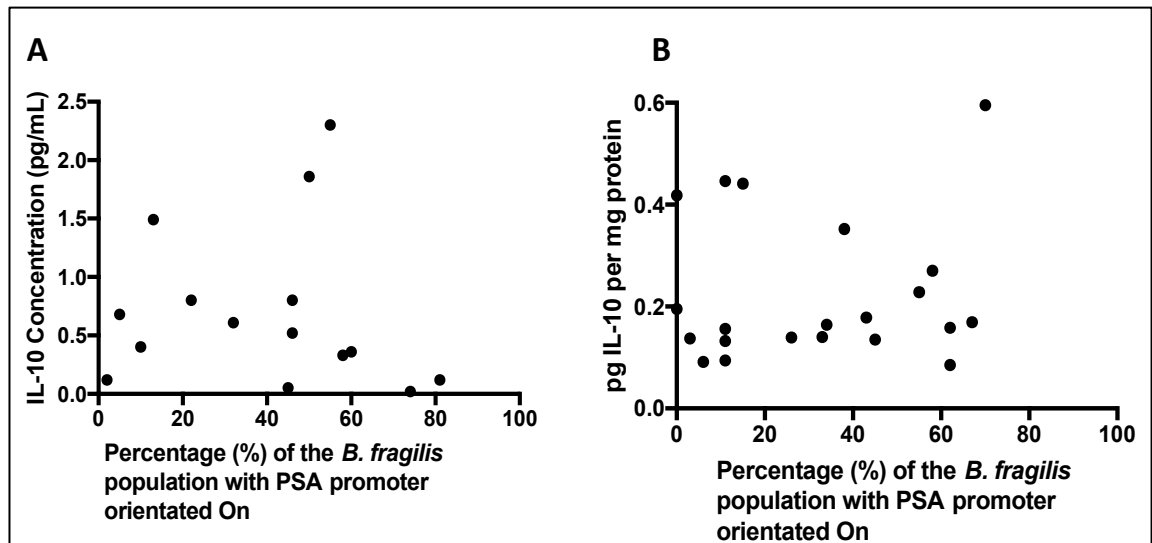
**Figure 5-14 Tissue IL-10 concentration as separated by the presence of BF or BFT**

Error bars for figures A and B indicate the median IL-10 concentration (pg/mL) and 95% confidence level (C.I.). A Mann-Whitney U test was used to determine any significant differences between positive and negative samples (Table 5-2)

(A) Serum IL-10 levels separated by individuals with biopsy samples positive for *B. fragilis*.

(B) Tissue IL-10 levels separated by individuals with biopsy samples positive for the *B. fragilis* toxin gene (*bft*). A Mann-Whitney U test was used to determine any significant differences between positive and negative samples.

## 5.4.3 Correlation analyses



**Figure 5-15** Correlation analysis of serum IL-10 or tissue IL-10 and *B. fragilis* PSA promoter orientation

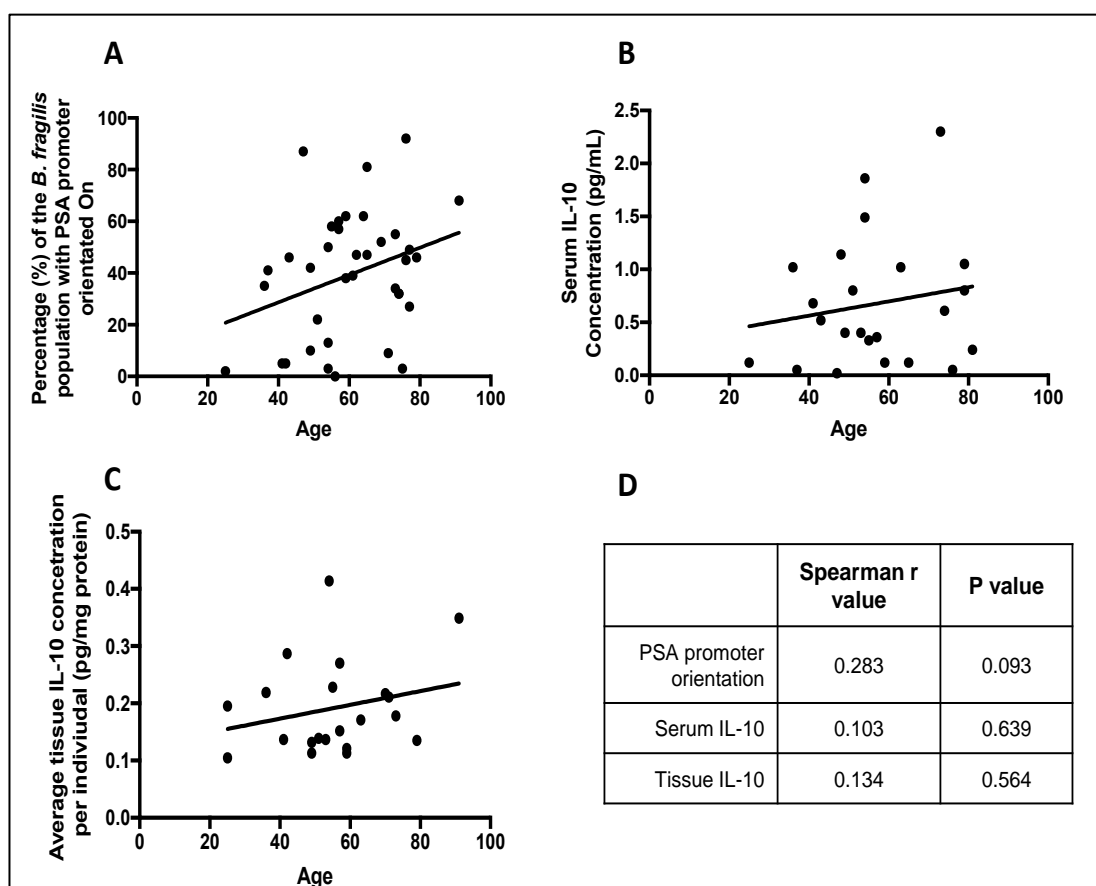
Correlation was tested statistically using Spearman's rank correlation

(A) Scatter graph to identify any correlation between serum IL-10 concentration (pg/mL) and *B. fragilis* PSA promoter orientation

(B) Scatter graph to identify any correlation between tissue IL-10 concentration (pg/mg protein) and *B. fragilis* PSA promoter orientation

The reasoning behind measuring IL-10 concentrations was to establish if they correlated with the percentage of *B. fragilis* cells with PSA promoter orientation. No linear relationship (Figure 5-15) was found with either serum IL-10 concentration (Spearman  $r = -0.225$ ,  $p = 0.417$ ) or biopsy tissue IL-10 (Spearman  $r = 0.134$ ,  $p = 0.564$ ). The product moment correlation coefficient ( $r$ ) for serum IL-10 is a negative value; therefore if any correlation exists between the two variables it is negative. The  $r$ -value is close to zero for biopsy tissue IL-10 suggesting there is very little or no correlation between the two variables. No correlation was also observed between tissue lysate IL-10 and serum IL-10 ( $p = 0.894$ ) and the Spearman  $r$  value was negative ( $r = -0.049$ ).

Age did not correlate with either serum (Figure 5-16B) or tissue IL-10 (Figure 1-16C) concentrations. A weak positive correlation ( $r = 0.283$ ) between age and the percentage of the *B. fragilis* population with the PSA promoter orientated was identified although this was not significant ( $p = 0.093$ ).



**Figure 5-16 Scatter graphs to identify possible correlation of variables with age of the individual**

(A) Scatter graph to identify any correlation between age and *B. fragilis* PSA orientation

(B) Scatter graph to identify any correlation between age and serum IL-10 concentration (pg/mL)

Error bars for figures A and B indicate the median IL-10 concentration (pg/mL) and 95% confidence level (C.I.).

(C) Error bars for figures A and B indicate the median IL-10 concentration (pg/mL) and 95% confidence level (C.I.).

Scatter graph to identify any correlation between age and tissue IL-10 concentration (pg/mg protein)

(D) Summary of Spearman's rank correlation used to statistically identify any correlation between age and all variables tested. Spearman's r values and p values are summarised in the table.

#### 5.4.4 Summary of Results

Presented in this section is a summary of the above results combining PSA promoter orientation and IL-10 levels. Table 5-2 shows that the only PCR gene target with a significant association with one of the aforementioned variables is *bft* and PSA promoter orientation, as described

previously (section 5.3.4). *B. fragilis* was close to being significantly associated with tissue IL-10 concentration and so may have some biological significance.

**Table 5-2 Summary of p values for significant differences between the variables tested (PSA promoter orientation, serum IL-10 concentration and tissue IL-10 concentration) and presence of bacterial taxa or virulence genes as detected by PCR**

PCR Gene Target	p (Mann-Whitney test)		
	PSA	serum IL-10	tissue IL-10
<i>B. fragilis</i> 16S rRNA	N/A	0.932	0.069
<i>bft</i>	0.000 ***	N/A	0.355
<i>Fusobacterium</i> 16S rRNA	0.520	0.839	0.710
<i>FadA</i>	>0.999	0.257	0.860

\*\*\* = p value of less than 0.001

Pairwise comparison of patient groups with PSA promoter orientation, serum IL-10 or tissue IL-10 (Table 5-3) highlights the significant associations between, IBD and the healthy group and the previous polyps/cancer cohort in respect to PSA promoter orientation (section 5.3.4).

**Table 5-3 Summary of p values for pairwise significant differences between patient groups for all of the variables tested (PSA promoter orientation, serum IL-10 concentration and tissue IL-10 concentration)**

Patient Groups	p (Kruskal-Wallis test)		
	PSA	serum IL-10	tissue IL-10
Healthy vs previous polyps/cancer	>0.999	>0.999	>0.999
Healthy vs new polyps/cancer	>0.999	>0.999	0.506
Healthy vs IBD	0.018*	>0.999	0.396
Previous polyps/cancer vs new polyp/cancer	0.969	>0.999	0.769
Previous polyps/cancer vs IBD	0.004**	>0.999	0.595
New polyps/cancer vs IBD	0.058	>0.999	>0.999

\* = p value of less than 0.05; \*\* = p value of less than 0.01

## 5.5 Discussion

### 5.5.1 PSA promoter orientation in vivo

The PSA promoter orientation in healthy individuals had a 95% C.I. around the median of between 40% and 60% of the *B. fragilis* population (Figure 5-8). PSA promoter orientation has not previously been studied in human tissue samples, with the nearest comparable research



#### Chapter 5.5.1 PSA promoter orientation in vivo

carried out in faecal samples from various mouse models (Troy *et al.*, 2010). They found that in monocolonised mice after 9 weeks approximately 80% of *B. fragilis* has the PSA promoter orientated 'ON', but, after 11 weeks, this value decreased towards ~50%. The results presented in this chapter suggest that in intestinal homeostasis and with *B. fragilis* in a commensal form there is a heterogenous population in terms of capsular polysaccharide.

When a complex microbiota from two different litters was used one litter had close to 80% of the population oriented on, while the other litter had close to 0%. This demonstrates how the microbiota composition may be an environmental factor for influencing PSA promoter orientation. Troy *et al.*, 2001 also found that broth cultures in exponential phase had a high percentage orientated on as did the first time point for monocolonised mice, so perhaps this a feature of an actively growing *B. fragilis* population. The results presented here are unlikely to be directly comparable to the study by Troy *et al.*, 2001 as the digestive system of mice and humans differ (Nguyen *et al.*, 2015). Mice have a comparatively larger caecum and colon than humans and are coprophagic, denoting that they consume each other's faeces, therefore within a litter the inter-mouse microbiota is likely to be consistent.

The results presented here show that in IBD patients *B. fragilis* populations have a significantly lower percentage of bacteria with the PSA promoter orientated in the 'on' position, although there is no significant difference based on sample location within an individual. This is an interesting result given that PSA can prevent *H. hepaticus*-induced colitis in mice (Mazmanian *et al.*, 2008). Previous research (Troy *et al.*, 2010) implicates microbiota composition as important for determining PSA promoter orientation and given that IBD individuals are known to have a dysbiotic microbiota this may be one explanation for my findings. Prevention of *H. hepaticus*-induced colitis is IL-10 mediated, but the results presented in this thesis show no significant difference in tissue IL-10 concentration between healthy individuals and those with IBD. Insufficient samples were tested to compare any potential correlation between PSA orientation and IL-10 concentration within the IBD patient group, although no correlation was observed when comparing all samples (Figure 5-15). A feature of IBD is a lack of immunotolerance towards commensal bacteria. The results reported in the current study may reflect the bacterium requiring phenotypic variation of capsular polysaccharides to evade the dysregulated immune response.

The only PCR target found to have a significant link with PSA promoter orientation was *bft*. Perhaps this association is observed because BFT is more likely to be found in individuals with GI disease, while PSA is more likely to be switched 'OFF' when intestinal homeostasis is disrupted i.e. in IBD. An alternative view is that BFT is a virulence factor associated with *B. fragilis* the pathogen and so will use capsular switching to avoid host adaptive immune responses or requires a new capsule for colonisation outside of the usual environmental niche.

### 5.5.2 Identification of a SNP

Whilst determining the *B. fragilis* PSA promoter orientation in patient biopsies an alternative cut pattern was observed leading to the identification of a SNP in the *upaY* gene. Due to the degeneracy of the genetic code, the DNA sequence change observed does not result in a difference in amino acid sequence. Serum and tissue IL-10 results were not available for all individuals with this SNP as it was not feasible to collect serum from every individual and so there were insufficient data to observe any downstream, functional consequences in IL-10 response of these individuals. Conversely, it has been found the *upaY* is conserved across many *Bacteroides* capsular polysaccharides (Krinos *et al.*, 2001) and in the above DNA sequence alignment (Figure 5-5) the SNP is surrounded by a conserved DNA sequence. This study found the SNP-containing strain in multiple individuals so perhaps there is an, as yet, unidentified reason for its positive selection. A possible result of the SNP, despite an unchanged protein sequence, could be an altered secondary structure at the 5' end of mRNA which is capable of influencing translational efficiency (de Smit and van Duin, 1990). As the UpxY family of proteins are involved in prevention of premature transcription termination within the 5' untranslated region between the promoter and the *UpxY* gene (Chatzidaki-Livanis *et al.*, 2009), and assuming the SNP results in reduced translational efficiency of *upaY* this would result in reduced expression of PSA.

### 5.5.3 Serum IL-10

The anti-inflammatory cytokine IL-10 is considered to have an important role in maintaining gastrointestinal homeostasis. Polymorphisms in the *IL-10* gene have been linked to IBD susceptibility (Tagore *et al.*, 1999) and a meta-analysis has shown that a high level of serum IL-10 is significantly associated with a worse survival outcome in cancer patients (Zhao *et al.*,

2015). This study found no significant difference in serum IL-10 concentrations amongst the different patient groups tested. Table 5-4 summarises literature measuring serum IL-10 concentrations in CRC and IBD groups. Mean serum concentrations in the healthy group varied across the studies, but a consistent finding across most of the literature is that healthy IL-10 levels are lower than in IBD or CRC cohorts. Furthermore, Stanilov *et al.*, 2010 show that increased IL-10 levels are associated with later stages of CRC. The results presented here are in agreement in that median serum IL-10 was higher in the tumour and IBD groups compared with healthy individuals, although not statistically significant.

**Table 5-4 Summary of literature measuring serum IL-10 Levels in IBD and CRC**

Authors	Method	Mean serum IL-10 levels (pg/mL)			
		CD	UC	CRC	Healthy
(Kucharzik <i>et al.</i> , 1995)	Sandwich ELISA developed in house	Active: 73 Inactive: 132	Active: 144 Inactive: 52	N/A	44
(O'Hara <i>et al.</i> , 1998)	ELISA (R&D systems)	N/A	N/A	Median: 4.11	Median: 2.56
(Szkardkiewicz <i>et al.</i> , 2009)	Quantikine ELISA (R&D Systems)	2.16	4.40	1.99	1.35
(Stanilov <i>et al.</i> , 2010)	IL-10 ELISA (Invitrogen)	N/A	N/A	16.09	5.1
(Wang <i>et al.</i> , 2011)	ELISA (BD)	4.77	N/A	N/A	2.58
(Miteva <i>et al.</i> , 2014)	ELISA (R&D systems)	N/A	N/A	12.6	6.42
(Abtahi <i>et al.</i> , 2017)	ELISA (eBioscience)	N/A	N/A	0.452	0.812

CD = Crohn's Disease; UC = Ulcerative colitis

#### 5.5.4 Tissue IL-10

Tissue IL-10 levels, similarly to serum IL-10 concentrations, showed no significant difference amongst patient groups. The lowest p value was observed comparing the healthy and IBD groups so perhaps a larger IBD cohort separating CD and UC patients would see significant differences. Tissue IL-10 levels are also going to depend on the vascularisation of the mucosal biopsy tissue, which may be more pronounced in tumour tissue.

No correlation was observed between tissue IL-10 concentration and serum IL-10 levels for individuals where both samples were measured. A possible explanation for this is that immune cell responses can be very tissue-specific (Hu and Pasare, 2013).

### **5.5.5 PSA promoter orientation is not correlated with IL-10 concentration**

No correlation was identified between PSA promoter orientation and either serum or tissue IL-10 levels (Figure 5-15). Our hypothesis that a positive correlation would be observed between the PSA promoter orientation and either serum or tissue IL-10 levels was not fulfilled. Possible reasons for why this may be the case are discussed below.

#### **5.5.5.1 Other zwitterionic capsular polysaccharides**

*B. fragilis* have been shown to synthesis chemically different versions of PSA, for example *B. fragilis* strain 638R expresses PSA2 (Wang *et al.*, 2000). The restriction digest method used in this thesis employed primers directed at inverted repeats surrounding the PSA operon and would not differentiate between these chemically different, and possibly dissimilar immune-stimulatory versions of PSA. *B. fragilis* is able to synthesise many different types of polysaccharide, including PSB that is also zwitterionic (Baumann *et al.*, 1992). Altogether this adds to the complexity of the relationship between PSA promoter orientation and IL-10 levels.

Other bacterial species have been shown to synthesise ZPSs. *Streptococcus pneumoniae* can synthesise the zwitterionic type 1 polysaccharide (Sp1) (Lindberg *et al.*, 1980) while another well studied zwitterionic capsule is *Staphylococcus aureus* CP8 (Wu and Park, 1971). Taxonomically diverse species, including other *Bacteroides* species, are predicted to produce the amino sugar acetamino-2, 4, 6-trideoxygalactose (AATGal). AATGal contains a positive charge key for the immunomodulatory properties of zwitterionic capsular polysaccharides (Neff *et al.*, 2016); therefore, it is proposed that taxonomically diverse gastrointestinal bacteria can produce immunomodulatory zwitterionic capsular polysaccharides. This would also explain the lack of correlation seen between PSA orientation and IL-10 concentrations, as the complexity of this potential correlation is beyond the scope of this study.

#### **5.5.6 The biological significance of capsular phase variation**

In the healthy cohort the PSA promoter is orientated on in approximately 50% of the *B. fragilis* population. This concurs with a theory of the biological significance of phase variation that it creates heterogeneity within a clonal population allowing for diversity to protect against sudden environmental stresses (Van Der Woude and Bäumlér, 2004). Other possible reasons for capsular polysaccharide phase variation include temporal heterogeneity that allows for

#### Chapter 5.5.7 Future research

avoidance of specific immune responses or colonisation of multiple niches within the host (Weiser *et al.*, 1994).

Evidence from *B. fragilis* research has shown that PSA can activate TLR2 on Fox3p+ regulatory T cells reducing Th17 responses (Round *et al.*, 2011). Suppression of Th17 by PSA allows for a small population of mucosal-associated *B. fragilis* to survive within colonic crypts. This may represent a small quantity of *B. fragilis*, but they are possibly more potent in their immunomodulatory capacity due to their expression of PSA. A stable acapsular mutant *B. fragilis* has been shown to be outcompeted by *B. fragilis* with the ability to synthesise one capsular polysaccharide (Coyne *et al.*, 2008), further suggesting the importance of the capsule for colonisation of the bacterium.

#### **5.5.7 Future research**

To begin to answer the biological significance of phase variable capsules in *B. fragilis*, future work should elucidate whether the difference observed in the IBD cohort is due to *B. fragilis* cells actively employing phase variation or PSA expressing populations being selected against based on host selection pressures.

Assuming promoter orientation relates to expression, the output of the restriction digest method used in this study describes the proportion of the *B. fragilis* population expressing PSA, but did not quantify the amount of *B. fragilis* and/or PSA. This could be significant as a threshold level may be required to stimulate immune cell populations. Future research in this area should also use immunohistochemistry or fluorescent in-situ hybridisation (FISH) to obtain localisation information about the bacterium, as this could be important for PSA-induced stimulation of different immune cell populations. This study found no significant difference in PSA promoter orientation within *B. fragilis* populations between different longitudinal colonic locations, but it is feasible that transverse localisation is more relevant, for example is *B. fragilis* lumenally located versus deep within colonic crypts. Perhaps a more appropriate method is required to look at the downstream consequence of an altered PSA promoter orientation in the IBD group. A more relevant output of PSA immunomodulation could be measurement of specific IL-10 responses in peripheral blood mononuclear cells (PBMCs), lamina propria mononuclear cells (LPMCs) or IL-

#### Chapter 5.5.7 Future research

10 activation of signalling pathways, possibly by using RT-qPCR to measure gene expression of IL-10 regulated genes.

A more direct method of measuring PSA expression such as an antibody to *B. fragilis* PSA or AATGal-ZPS would bypass any mechanisms of transcriptional regulation. The field requires more knowledge on expression conditions and mechanisms of the various *B. fragilis* capsules and following the findings presented here particularly how this relates to IBD. Future work could investigate PSA promoter orientation in a larger cohort of individuals with IBD with the aim of investigating differences between active and inactive disease and between CD and UC. Finally, a more speculative approach could be the use of PSA promoter orientation as a putative biomarker for gastrointestinal health; however more research is required first to see whether faecal samples compare to mucosal biopsies and there is a limitation in that *B. fragilis* is not ubiquitous.

## Chapter 6 : GENERAL DISCUSSION

Presented in this thesis is a PCR-based method for the detection of bacterial taxa and genes hypothesised to be important in gastrointestinal health. The gene coding for a putative oncogenic *B. fragilis* virulence factor, *bft*, was found at a significantly higher prevalence in individuals newly diagnosed with polyps/cancer compared with a healthy cohort. Prevalences of *Fusobacterium* and *fadA* were not significantly different in the cohorts investigated in this study. Detection rates of *bft* and *B. fragilis ubb* were higher in IBD patients compared with healthy individuals. These last results were not significant but merit further study.

Colonic location and histological type of *Fusobacterium*-positive tumours did not exhibit any significant associations, but the trends observed support previous suggestions of an association between *Fusobacterium* species and right-sided colon cancer (Mima, 2016). Similarly, no statistical significance between *bft* and tumour location was observed but the results support the association of *bft* with left-sided tumours (Purcell *et al.*, 2017). In addition, a potential feature of *bft* tumorigenesis is the development of multiple polyps and the growth of new tumours following polypectomy. This observation was not significant in this study and therefore requires further research involving a larger sample size.

This thesis also contains the first reported determination of *B. fragilis* capsular PSA promoter orientation *in vivo*. In a healthy population the PSA promoter was on in 40-60% of individuals, whereas a significantly lower percentage of the *B. fragilis* population had PSA orientated on in individuals with IBD and positive for *bft*. PSA promoter orientation may therefore signify a novel biomarker of gastrointestinal health. One possible consequence of decreased PSA expression is reduced IL-10 production, but this was not observed in this study when looking at serum or tissue levels.

CRC has been classified into different numbers of subtypes using a variety of methods, including whole genome sequencing (Muzny *et al.*, 2012), gene expression (Guinney *et al.*, 2015) and proteomics (Zhang *et al.*, 2014). The results from this thesis did not find *Fusobacterium* and *bft* in all polyps and cancer samples supporting the idea that a subgroup of CRC could be microbial-induced, or that CRC categories with different properties are a result of

### Chapter 6.1.1 Polyposis

diverse oncogenic microbes. It should not be forgotten that colon cancer develops over many years and therefore the tumour environmental niche, and the tumour microbiome, could change temporally. The results presented in this thesis do not disprove or support the driver-passenger model (section 1.7.1) but have shown that *FadA* positive *Fusobacterium* species can coexist in cancer tissue with *bft* and so do not have to be mutually exclusive (Tjalsma *et al.*, 2012).

#### 6.1.1 Polyposis

A higher prevalence of *bft* was found in individuals with more than one polyp/cancer compared with patients with a single colonic tumour. This finding was not significant, which may reflect the small sample size. This is believed to be the first study to compare single and synchronous polyps in terms of the microbiome. The microbiome of patients with the inherited polyposis syndrome FAP (section 1.4.4) has been investigated, although this was in respect to dysbiosis in pouchitis (Tannock *et al.*, 2012) (Zella *et al.*, 2011). A study found that 23.5% of patients with CRC have synchronous polyps (Demetriades *et al.*, 2004), raising the question of whether there is a different etiology in patients with synchronous tumours compared with a single tumour? A meta-analysis (Jayasekara *et al.*, 2017) showed that a proximal location of CRC is a risk factor for metachronous CRC, likewise *Fusobacterium* is associated with the right colon (Mima *et al.*, 2016, Yu *et al.*, 2016).

Similarly, when comparing individuals with a history of polyps/cancer the detection rate of *bft* was higher in individuals with new tumours compared with a clear colonoscopy outcome. As with multiple polyps, it appears that these groups have not previously been compared in respect to the microbiome.

#### 6.1.2 Biogeography

Differences in prevalence results between caecal effluent (CE) and biopsy samples from the same individual could be for a number of reasons. The CE sample was analysed by a culture-dependent method in comparison to the culture-independent DNA extract from biopsy tissue. Future work would use the DNA extraction method on CE to compare with the culturing results. In addition, the culturing study would be extended to all individuals from which caecal effluent



was collected. This would give a clearer picture as to whether the results are dependent on the type of methodology or the type of sample and its location within the colon.

Just one study has associated *bft* with tumour location and histological type (Purcell *et al.*, 2017). ETBF positivity was found to be associated with left-sided tumours; similarly, out 5 *bft*-positive individuals with tumours presented in this thesis 3 were sigmoidal and 1 was in the descending colon. Purcell *et al.*, 2017 also found an association between *bft* and serrated polyps and tubular adenomas. My results support this finding as 3 out of 5 individuals positive for *bft* had either sessile, serrated polyps or tubular adenomas. The other *bft*-positive tumours were rare inflammatory polyps not present in the Purcell *et al.*, 2017 study.

Future research in this field requires more information on localization and biogeography of the different taxonomic groups. The literature has consistently found *Fusobacterium* to be more abundant on tumour tissue compared with adjacent healthy mucosa (Tahara *et al.*, 2015) but a study of the oral microbiome using FISH (Mark Welch *et al.*, 2016) showed a distinct spatial arrangement of the microbiome with less abundant species in potentially important geographical patterns. Blind sampling of the oral microbiome without this biogeographical knowledge could result in abundance-based results misinterpreting biological relevance. In respect to the colorectal tumour microbiome it is therefore important to first establish what is present, then to answer where is it, and finally what is being expressed.

### **6.1.3 *Fusobacterium* and *F. nucleatum* taxonomy**

Comparative genomics of whole bacterial genomes revealed *Fusobacterium* species fall into three main lineages (McGuire *et al.*, 2014). Research into *F. nucleatum* as a pathogen often overlooks the heterogeneity between the different subspecies. Multiple studies have suggested *F. nucleatum* subspecies are diverse enough to be considered separate species (Manson McGuire *et al.*, 2014, Kook *et al.*, 2017). The expanded HMP dataset (HMP1-II) revealed that *F. nucleatum* is niche-associated for body site (Lloyd-Price *et al.*, 2017) suggesting that oral and gastrointestinal strains may represent different subspecies. Phylogenetic analysis using the RNA polymerase  $\beta$  subunit (*rpoB*) gene and the 16S rRNA gene revealed a distinction between oral and gastrointestinal isolates (Strauss *et al.*, 2008). Strauss *et al.*, 2008 found subspecies *animalis* to be the most common type isolated from gut biopsies, while this subspecies has

rarely been found in the oral cavity (Gmur *et al.*, 2006). *F. nucleatum* subsp. *animalis* has also been found to be the most prevalent subspecies in CRC (Ye *et al.*, 2017). *F. nucleatum* strains can differ in their invasiveness, a feature particularly associated with IBD (Strauss *et al.*, 2011), with an especially invasive isolate shown to have bacteriophage within its genome (Cochrane *et al.*, 2016). Strain importance was also highlighted in the Apc<sup>(Min/+)</sup> mouse model of intestinal tumorigenesis, where one strain initiated tumorigenesis (Kostic *et al.*, 2013) and another did not replicate this finding (Tomkovich *et al.*, 2017). Altogether the literature on *F. nucleatum* points towards specific strains, and perhaps even particular virulence genes, being associated with IBD/CRC.

Whole genome sequencing of *Ruminococcus gnavus* strains isolated from healthy and IBD patients revealed another example of the importance of strain-level information (Hall *et al.*, 2017). Two clades of *R. gnavus* were identified with one of the strains enriched in individuals with IBD; consequently some of these bacterial genes were almost exclusively present in IBD patients.

In conclusion, overall the results presented indicate that the common gastrointestinal species, *B. fragilis*, can have wide ranging effects on gastrointestinal health. Between-strain differences and within-strain antigenic variation were shown to have significant associations with patient populations and argue for a gene-centric, and not taxonomic-centric, approach to microbiome research.

## REFERENCES

- ABNET, C. C., et al.** (2001) Prospective study of tooth loss and incident esophageal and gastric cancers in China. *Cancer Causes Control*, **12** (9), 847-54.
- ABTAHI, S., et al.** (2017) Dual association of serum interleukin-10 levels with colorectal cancer. *J Cancer Res Ther*, **13** (2), 252-256.
- AHMED, A. U.** (2011) An overview of inflammation: mechanism and consequences. *Frontiers in Biology*, **6** (4), 274-281.
- AITCHISON, A., FRIZELLE, F. A. & KEENAN, J. I.** (2016) PCR Detection of the *Bacteroides fragilis* Enterotoxin Gene Relies on Robust Primer Design. *J Clin Microbiol*, **54** (1), 239-40.
- ALLEN-VERCOE, E., STRAUSS, J. & CHADEE, K.** (2011) *Fusobacterium nucleatum*: an emerging gut pathogen? *Gut Microbes*, **2** (5), 294-8.
- AL-MALEKI, A. R., et al.** (2017) *Helicobacter pylori* outer inflammatory protein A (OipA) suppresses apoptosis of AGS gastric cells in vitro.) *Cellular Microbiology*.
- AMIN, M. B., AMERICAN JOINT COMMITTEE ON CANCER. & AMERICAN CANCER SOCIETY.** (2017) *AJCC cancer staging manual*. Chicago IL: American Joint Committee on Cancer, Springer.
- ANDREWES, F. W.** (1922) Studies in group-agglutination I The salmonella group and its antigenic structure. *Journal of Pathology and Bacteriology*, **25** (4), 505-521.
- AVILA-CAMPOS, M. J., LIU, C., SONG, Y., ROWLINSON, M. C. & FINEGOLD, S. M.** (2007) Determination of *bft* gene subtypes in *Bacteroides fragilis* clinical isolates. *J Clin Microbiol*, **45** (4), 1336-8.
- BASSET, C., HOLTON, J., BAZEOS, A., VAIRA, D. & BLOOM, S.** (2004) Are *Helicobacter* species and enterotoxigenic *Bacteroides fragilis* involved in inflammatory bowel disease? *Dig Dis Sci*, **49** (9), 1425-32.
- BAUMANN, H., TZIANABOS, A. O., BRISSON, J. R., KASPER, D. L. & JENNINGS, H. J.** (1992) Structural elucidation of two capsular polysaccharides from one strain of *Bacteroides fragilis* using high-resolution NMR spectroscopy. *Biochemistry*, **31** (16), 4081-9.
- BAYLISS, R., et al.** (1984) The bowel, the genitourinary tract, and infective endocarditis. *Br Heart J*, **51** (3), 339-45.
- BEAUDOIN, M., et al.** (2013) Deep resequencing of GWAS loci identifies rare variants in CARD9, IL23R and RNF186 that are associated with ulcerative colitis. *PLoS Genet*, **9** (9), e1003723.
- BEECHING, N. J., CHRISTMAS, T. I., ELLIS-PEGLER, R. B. & NICHOLSON, G. I.** (1985) *Streptococcus bovis* bacteraemia requires rigorous exclusion of colonic neoplasia and endocarditis. *Q J Med*, **56** (220), 439-50.
- BENSON, D. A., et al.** (2013) GenBank. *Nucleic Acids Res*, **41** (Database issue), D36-42.
- BERGMANN, H., et al.** (2017) Card9-dependent IL-1 $\beta$  regulates IL-22 production from group 3 innate lymphoid cells and promotes colitis-associated cancer. *Eur J Immunol*, **47** (8), 1342-1353.
- BERGOUNIOUX, J., et al.** (2012) Calpain activation by the *Shigella flexneri* effector VirA regulates key steps in the formation and life of the bacterium's epithelial niche. *Cell Host Microbe*, **11** (3), 240-52.
- BIELASZEWSKA, M., SINHA, B., KUCZIUS, T. & KARCH, H.** (2005) Cytolethal distending toxin from Shiga toxin-producing *Escherichia coli* O157 causes irreversible G2/M arrest, inhibition of proliferation, and death of human endothelial cells. *Infect Immun*, **73** (1), 552-62.
- BLOOM, S. M., et al.** (2011) Commensal *Bacteroides* species induce colitis in host-genotype-specific fashion in a mouse model of inflammatory bowel disease. *Cell Host Microbe*, **9** (5), 390-403.
- BOLAND, C. R. and GOEL, A.** (2010) Microsatellite instability in colorectal cancer. *Gastroenterology*, **138** (6), 2073-2087 e3.
- BOLEIJ, A., et al.** (2015) The *Bacteroides fragilis* toxin gene is prevalent in the colon mucosa of colorectal cancer patients. *Clin Infect Dis*, **60** (2), 208-15.
- BOLEIJ, A., et al.** (2011) Novel clues on the specific association of *Streptococcus gallolyticus* subsp. *gallolyticus* with colorectal cancer. *J Infect Dis*, **203** (8), 1101-9.
- BOLEIJ, A., VAN GELDER, M. M., SWINKELS, D. W. & TJALSMA, H.** (2011) Clinical Importance of *Streptococcus gallolyticus* infection among colorectal cancer patients: systematic review and meta-analysis. *Clin Infect Dis*, **53** (9), 870-8.
- BONNET, M., et al.** (2014) Colonization of the human gut by *E. coli* and colorectal cancer risk. *Clin Cancer Res*, **20** (4), 859-67.
- BOSMAN, F. T., CARNEIRO, F., HRUBAN, R. H. & THEISE, N. D.** (2010) WHO Classification of Tumours of the Digestive System, Fourth Edition. In Vol. 3. pp. 417-417.
- BOVERI, T.** (2008) Concerning the origin of malignant tumours by Theodor Boveri. Translated and annotated by Henry Harris. *J Cell Sci*, **121 Suppl 1** (Supplement 1), 1-84.
- BROGDEN, K. A., et al.** (2000) *Virulence mechanisms of bacterial pathogens*. Washington, D.C: ASM Press.
- BRONTE-TINKEW, D. M., et al.** (2009) *Helicobacter pylori* cytotoxin-associated gene A activates the signal transducer and activator of transcription 3 pathway *in vitro* and *in vivo*. *Cancer Res*, **69** (2), 632-9.

- BROUWER, M. S., et al.** (2013) Horizontal gene transfer converts non-toxicogenic *Clostridium difficile* strains into toxin producers. *Nat Commun*, **4**, 2601.
- BROWN, S. P., CORNFORTH, D. M. & MIDEO, N.** (2012) Evolution of virulence in opportunistic pathogens: generalism, plasticity, and control. *Trends Microbiol*, **20** (7), 336-42.
- BUC, E., et al.** (2013) High prevalence of mucosa-associated *E. coli* producing cyclomodulin and genotoxin in colon cancer. *PLoS One*, **8** (2), e56964.
- BUCKWOLD, S. L., SHOEMAKER, N. B., SEARS, C. L. & FRANCO, A. A.** (2007) Identification and characterization of conjugative transposons CTn86 and CTn9343 in *Bacteroides fragilis* strains. *Appl Environ Microbiol*, **73** (1), 53-63.
- BUDA, A., et al.** (2014) Confocal laser endomicroscopy for prediction of disease relapse in ulcerative colitis: a pilot study. *J Crohns Colitis*, **8** (4), 304-11.
- BULLMAN, S., et al.** (2017) Analysis of *Fusobacterium* persistence and antibiotic response in colorectal cancer. *Science*, **358** (6369), 1443-1448.
- BUTI, L., et al.** (2011) *Helicobacter pylori* cytotoxin-associated gene A (CagA) subverts the apoptosis-stimulating protein of p53 (ASPP2) tumor suppressor pathway of the host. *Proc Natl Acad Sci U S A*, **108** (22), 9238-43.
- BUTT, J., et al.** (2016) Association of *Streptococcus gallolyticus* subspecies *gallolyticus* with colorectal cancer: Serological evidence. *Int J Cancer*, **138** (7), 1670-9.
- CASTELLARIN, M., et al.** (2012) *Fusobacterium nucleatum* infection is prevalent in human colorectal carcinoma. *Genome Res*, **22** (2), 299-306.
- CERDENO-TARRAGA, A. M., et al.** (2005) Extensive DNA inversions in the *B. fragilis* genome control variable gene expression. *Science*, **307** (5714), 1463-5.
- CHANG, J., et al.** (2017) Impaired Intestinal Permeability Contributes to Ongoing Bowel Symptoms in Patients With Inflammatory Bowel Disease and Mucosal Healing. *Gastroenterology*, **153** (3), 723-731 e1.
- CHATZIDAKI-LIVANIS, M., COYNE, M. J. & COMSTOCK, L. E.** (2009) A family of transcriptional antitermination factors necessary for synthesis of the capsular polysaccharides of *Bacteroides fragilis*. *J Bacteriol*, **191** (23), 7288-95.
- CHATZIDAKI-LIVANIS, M., et al.** (2017) Gut Symbiont *Bacteroides fragilis* Secretes a Eukaryotic-Like Ubiquitin Protein That Mediates Intraspecies Antagonism. *MBio*, **8** (6), 1-12.
- CHATZIDAKI-LIVANIS, M., GEVA-ZATORSKY, N. & COMSTOCK, L. E.** (2016) *Bacteroides fragilis* type VI secretion systems use novel effector and immunity proteins to antagonize human gut *Bacteroidales* species. *Proc Natl Acad Sci U S A*, **113** (13), 3627-32.
- CHATZIDAKI-LIVANIS, M., WEINACHT, K. G. & COMSTOCK, L. E.** (2010) Trans locus inhibitors limit concomitant polysaccharide synthesis in the human gut symbiont *Bacteroides fragilis*. *Proc Natl Acad Sci U S A*, **107** (26), 11976-80.
- CHEN, Y., et al.** (2017) Invasive *Fusobacterium nucleatum* activates beta-catenin signaling in colorectal cancer via a TLR4/P-PAK1 cascade. *Oncotarget*, **8** (19), 31802-31814.
- CHENG, F., et al.** (2003) A critical role for Stat3 signaling in immune tolerance. *Immunity*, **19** (3), 425-36.
- CHUNG, G. T., et al.** (1999) Identification of a third metalloprotease toxin gene in extraintestinal isolates of *Bacteroides fragilis*. *Infect Immun*, **67** (9), 4945-9.
- CLEVERS, H.** (2006) Wnt/beta-catenin signaling in development and disease. *Cell*, **127** (3), 469-80.
- COCHRANE, K., et al.** (2016) Complete genome sequences and analysis of the *Fusobacterium nucleatum* subspecies *animalis* 7-1 bacteriophage Funu1 and Funu2. *Anaerobe*, **38**, 125-129.
- COMSTOCK, L. E., et al.** (1999) Analysis of a capsular polysaccharide biosynthesis locus of *Bacteroides fragilis*. *Infect Immun*, **67** (7), 3525-32.
- CONE, R. A.** (2009) Barrier properties of mucus. *Adv Drug Deliv Rev*, **61** (2), 75-85.
- COONEY, R., et al.** (2010) NOD2 stimulation induces autophagy in dendritic cells influencing bacterial handling and antigen presentation. *Nat Med*, **16** (1), 90-7.
- COUGNOUX, A., et al.** (2014) Bacterial genotoxin colibactin promotes colon tumour growth by inducing a senescence-associated secretory phenotype. *Gut*, **63** (12), 1932-42.
- COYNE, M. J., CHATZIDAKI-LIVANIS, M., PAOLETTI, L. C. & COMSTOCK, L. E.** (2008) Role of glycan synthesis in colonization of the mammalian gut by the bacterial symbiont *Bacteroides fragilis*. *Proc Natl Acad Sci U S A*, **105** (35), 13099-104.
- COYNE, M. J., WEINACHT, K. G., KRINOS, C. M. & COMSTOCK, L. E.** (2003) Mpi recombinase globally modulates the surface architecture of a human commensal bacterium. *Proc Natl Acad Sci U S A*, **100** (18), 10446-51.
- CUEVAS-RAMOS, G., et al.** (2010) *Escherichia coli* induces DNA damage in vivo and triggers genomic instability in mammalian cells. *Proc Natl Acad Sci U S A*, **107** (25), 11537-42.
- CUI, J., et al.** (2010) Glutamine deamidation and dysfunction of ubiquitin/NEDD8 induced by a bacterial effector family. *Science*, **329** (5996), 1215-8.

- DASGUPTA, S., ERTURK-HASDEMIR, D., OCHOA-REPARAZ, J., REINECKER, H. C. & KASPER, D. L.** (2014) Plasmacytoid dendritic cells mediate anti-inflammatory responses to a gut commensal molecule via both innate and adaptive mechanisms. *Cell Host Microbe*, **15** (4), 413-23.
- DE RYCKE, J., et al.** (1997) Enteropathogenic *Escherichia coli* O103 from rabbit elicits actin stress fibers and focal adhesions in HeLa epithelial cells, cytopathic effects that are linked to an analog of the locus of enterocyte effacement. *Infect Immun*, **65** (7), 2555-63.
- DE SMIT, M. H. and VAN DUIN, J.** (1990) Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proc Natl Acad Sci U S A*, **87** (19), 7668-72.
- DEBONGNIE, J. C. and PHILLIPS, S. F.** (1978) Capacity of the human colon to absorb fluid. *Gastroenterology*, **74** (4), 698-703.
- DEFILIPPIS, R. A., GOODWIN, E. C., WU, L. & DIMAIO, D.** (2003) Endogenous human papillomavirus E6 and E7 proteins differentially regulate proliferation, senescence, and apoptosis in HeLa cervical carcinoma cells. *J Virol*, **77** (2), 1551-63.
- DEMETRIADES, H., et al.** (2004) Synchronous polyps in patients with colorectal cancer. *Tech Coloproctol*, **8** Suppl 1 (SUPPL. 1), s72-5.
- DEN BESTEN, G., et al.** (2013) The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *J Lipid Res*, **54** (9), 2325-40.
- DJEKIC, A. and MULLER, A.** (2016) The Immunomodulator VacA Promotes Immune Tolerance and Persistent *Helicobacter pylori* Infection through Its Activities on T-Cells and Antigen-Presenting Cells. *Toxins (Basel)*, **8** (6).
- DONNELLY, R. P., DICKENSHEETS, H. & FINBLOOM, D. S.** (1999) The interleukin-10 signal transduction pathway and regulation of gene expression in mononuclear phagocytes. *J Interferon Cytokine Res*, **19** (6), 563-73.
- DOUGLAS, J. A., et al.** (2005) History and molecular genetics of Lynch syndrome in family G: a century later. *JAMA*, **294** (17), 2195-202.
- DREWES, J. L., et al.** (2017) High-resolution bacterial 16S rRNA gene profile meta-analysis and biofilm status reveal common colorectal cancer consortia. *NPJ Biofilms Microbiomes*, **3** (1), 34.
- DROLET, M., et al.** (2015) Population-level impact and herd effects following human papillomavirus vaccination programmes: a systematic review and meta-analysis. *Lancet Infect Dis*, **15** (5), 565-80.
- DUAN, J., AVCI, F. Y. & KASPER, D. L.** (2008) Microbial carbohydrate depolymerization by antigen-presenting cells: deamination prior to presentation by the MHCII pathway. *Proc Natl Acad Sci U S A*, **105** (13), 5183-8.
- DZINK, J. L., SHEENAN, M. T. & SOCRANSKY, S. S.** (1990) Proposal of three subspecies of *Fusobacterium nucleatum* Knorr 1922: *Fusobacterium nucleatum* subsp. *nucleatum* subsp. nov., comb. nov.; *Fusobacterium nucleatum* subsp. *polymorphum* subsp. nov., nom. rev., comb. nov.; and *Fusobacterium nucleatum* subsp. *vincentii* subsp. nov., nom. rev., comb. nov. *Int J Syst Bacteriol*, **40** (1), 74-8.
- ELLIS, U.** (2010) Mosby's medical dictionary. 8th ed.).
- FAHRER, J., et al.** (2014) Cytolethal distending toxin (CDT) is a radiomimetic agent and induces persistent levels of DNA double-strand breaks in human fibroblasts. *DNA Repair (Amst)*, **18**, 31-43.
- FAKHOURY, M., NEGRULJ, R., MOORANIAN, A. & AL-SALAMI, H.** (2014) Inflammatory bowel disease: clinical aspects and treatments. *J Inflamm Res*, **7**, 113-20.
- FEDOR, Y., et al.** (2013) From single-strand breaks to double-strand breaks during S-phase: a new mode of action of the *Escherichia coli* Cytolethal Distending Toxin. *Cell Microbiol*, **15** (1), 1-15.
- FELSENSTEIN, J.** (1985) Confidence-Limits on Phylogenies - an Approach Using the Bootstrap. *Evolution*, **39** (4), 783-791.
- FIORENTINI, C., et al.** (1988) Cytoskeletal changes induced in HEp-2 cells by the cytotoxic necrotizing factor of *Escherichia coli*. *Toxicon*, **26** (11), 1047-56.
- FRANCO, A. A.** (2004) The *Bacteroides fragilis* pathogenicity island is contained in a putative novel conjugative transposon. *J Bacteriol*, **186** (18), 6077-92.
- FRIES, W., et al.** (2005) Intestinal permeability and genetic determinants in patients, first-degree relatives, and controls in a high-incidence area of Crohn's disease in Southern Italy. *Am J Gastroenterol*, **100** (12), 2730-6.
- GAGNAIRE, A., NADEL, B., RAOULT, D., NEEFJES, J. & GORVEL, J. P.** (2017) Collateral damage: insights into bacterial mechanisms that predispose host cells to cancer. *Nat Rev Microbiol*, **15** (2), 109-128.
- GAGNIERE, J., et al.** (2017) Interactions between microsatellite instability and human gut colonization by *Escherichia coli* in colorectal cancer. *Clin Sci (Lond)*, **131** (6), 471-485.
- GALMICHE, A., et al.** (2000) The N-terminal 34 kDa fragment of *Helicobacter pylori* vacuolating cytotoxin targets mitochondria and induces cytochrome c release. *EMBO J*, **19** (23), 6361-70.
- GARROD, D. and CHIDGEY, M.** (2008) Desmosome structure, composition and function. *Biochim Biophys Acta*, **1778** (3), 572-87.

- GASSLER, N., et al.** (2001) Inflammatory bowel disease is associated with changes of enterocytic junctions. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, **281** (1), G216-G228.
- GEBOES, K.** (2003) Histopathology of Crohn's Disease and Ulcerative Colitis. *Inflammatory Bowel Disease*, **18**, 255-276.
- GHARBIA, S. E. and SHAH, H. N.** (1992) *Fusobacterium nucleatum* subsp. *fusiforme* subsp. nov. and *Fusobacterium nucleatum* subsp. *animalis* subsp. nov. as additional subspecies within *Fusobacterium nucleatum*. *Int J Syst Bacteriol*, **42** (2), 296-8.
- GLOCKER, E. O., et al.** (2009) Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. *N Engl J Med*, **361** (21), 2033-45.
- GMUR, R., MUNSON, M. A. & WADE, W. G.** (2006) Genotypic and phenotypic characterization of *fusobacteria* from Chinese and European patients with inflammatory periodontal diseases. *Syst Appl Microbiol*, **29** (2), 120-30.
- GOODWIN, A. C., et al.** (2011) Polyamine catabolism contributes to enterotoxigenic *Bacteroides fragilis*-induced colon tumorigenesis. *Proc Natl Acad Sci U S A*, **108** (37), 15354-9.
- GOULAS, T., AROLAS, J. L. & GOMIS-RUTH, F. X.** (2011) Structure, function and latency regulation of a bacterial enterotoxin potentially derived from a mammalian adamalysin/ADAM xenolog. *Proc Natl Acad Sci U S A*, **108** (5), 1856-61.
- GRILL, J. I., NEUMANN, J., HILTWEIN, F., KOLLIGS, F. T. & SCHNEIDER, M. R.** (2015) Intestinal E-cadherin Deficiency Aggravates Dextran Sodium Sulfate-Induced Colitis. *Dig Dis Sci*, **60** (4), 895-902.
- GROSS, O., et al.** (2006) Card9 controls a non-TLR signalling pathway for innate anti-fungal immunity. *Nature*, **442** (7103), 651-6.
- GROUP, N. H. W., et al.** (2009) The NIH Human Microbiome Project. *Genome Res*, **19** (12), 2317-23.
- GUIDI, R., et al.** (2013) Chronic exposure to the cytolethal distending toxins of Gram-negative bacteria promotes genomic instability and altered DNA damage response. *Cell Microbiol*, **15** (1), 98-113.
- GUINDI, M. and RIDDELL, R. H.** (2004) Indeterminate colitis. *J Clin Pathol*, **57** (12), 1233-44.
- GUINNEY, J., et al.** (2015) The consensus molecular subtypes of colorectal cancer. *Nat Med*, **21** (11), 1350-6.
- GUR, C., et al.** (2015) Binding of the Fap2 protein of *Fusobacterium nucleatum* to human inhibitory receptor TIGIT protects tumors from immune cell attack. *Immunity*, **42** (2), 344-355.
- HALF, E., BERCOVICH, D. & ROZEN, P.** (2009) Familial adenomatous polyposis. *Orphanet J Rare Dis*, **4**, 22.
- HALL, A. B., et al.** (2017) A novel *Ruminococcus gnavus* clade enriched in inflammatory bowel disease patients. *Genome Med*, **9** (1), 103.
- HAMPE, J., et al.** (2007) A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat Genet*, **39** (2), 207-11.
- HANAHAN, D. and WEINBERG, R. A.** (2000) The hallmarks of cancer. *Cell*, **100** (1), 57-70.
- HANAHAN, D. and WEINBERG, R. A.** (2011) Hallmarks of cancer: the next generation. *Cell*, **144** (5), 646-74.
- HARMON, B., et al.** (2016) A Genome-Wide RNA Interference Screen Identifies a Role for Wnt/beta-Catenin Signaling during Rift Valley Fever Virus Infection. *J Virol*, **90** (16), 7084-97.
- HATTA, K., NOSE, A., NAGAFUCHI, A. & TAKEICHI, M.** (1988) Cloning and expression of cDNA encoding a neural calcium-dependent cell adhesion molecule: its identity in the cadherin gene family. *J Cell Biol*, **106** (3), 873-81.
- HAZEWINKEL, Y. and DEKKER, E.** (2011) Colonoscopy: basic principles and novel techniques. *Nat Rev Gastroenterol Hepatol*, **8** (10), 554-64.
- HECHT, A. L., et al.** (2016) Strain competition restricts colonization of an enteric pathogen and prevents colitis. *EMBO Rep*, **17** (9), 1281-91.
- HIGASHI, H., et al.** (2002) Biological activity of the *Helicobacter pylori* virulence factor CagA is determined by variation in the tyrosine phosphorylation sites. *Proc Natl Acad Sci U S A*, **99** (22), 14428-33.
- HIGASHI, H., et al.** (2002) SHP-2 tyrosine phosphatase as an intracellular target of *Helicobacter pylori* CagA protein. *Science*, **295** (5555), 683-6.
- HORRIDGE, D. N., et al.** (2017) Outer inflammatory protein a (OipA) of *Helicobacter pylori* is regulated by host cell contact and mediates CagA translocation and interleukin-8 response only in the presence of a functional cag pathogenicity island type IV secretion system. *Pathog Dis*, **75** (8).
- HOUSSEAU, F. and SEARS, C. L.** (2010) Enterotoxigenic *Bacteroides fragilis* (ETBF)-mediated colitis in Min (Apc(+/-)) mice A human commensal-based murine model of colon carcinogenesis. *Cell Cycle*, **9** (1), 3-5.
- HOY, B., et al.** (2010) *Helicobacter pylori* HtrA is a new secreted virulence factor that cleaves E-cadherin to disrupt intercellular adhesion. *Embo Reports*, **11** (10), 798-804.
- HSU, Y. M., et al.** (2007) The adaptor protein CARD9 is required for innate immune responses to intracellular pathogens. *Nat Immunol*, **8** (2), 198-205.

- HU, W. and PASARE, C.** (2013) Location, location, location: tissue-specific regulation of immune responses. *J Leukoc Biol*, **94** (3), 409-21.
- HUANG, J. Y., LEE, S. M. & MAZMANIAN, S. K.** (2011) The human commensal *Bacteroides fragilis* binds intestinal mucin. *Anaerobe*, **17** (4), 137-41.
- HUGOT, J. P., et al.** (2001) Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature*, **411** (6837), 599-603.
- HUYCKE, M. M. and MOORE, D. R.** (2002) *In vivo* production of hydroxyl radical by *Enterococcus faecalis* colonizing the intestinal tract using aromatic hydroxylation. *Free Radical Biology and Medicine*, **33** (6), 818-826.
- HVEEM, T. S., et al.** (2014) Prognostic impact of genomic instability in colorectal cancer. *Br J Cancer*, **110** (8), 2159-64.
- HWANG, S., GWON, S. Y., KIM, M. S., LEE, S. & RHEE, K. J.** (2013) *Bacteroides fragilis* Toxin Induces IL-8 Secretion in HT29/C1 Cells through Disruption of E-cadherin Junctions. *Immune Netw*, **13** (5), 213-7.
- HYUN, J., et al.** (2015) Human intestinal epithelial cells express interleukin-10 through Toll-like receptor 4-mediated epithelial-macrophage crosstalk. *J Innate Immun*, **7** (1), 87-101.
- ISHIKAWA, K., ISHII, H. & SAITO, T.** (2006) DNA damage-dependent cell cycle checkpoints and genomic stability. *DNA Cell Biol*, **25** (7), 406-11.
- IWAMOTO, H., CZAJKOWSKY, D. M., COVER, T. L., SZABO, G. & SHAO, Z.** (1999) VacA from *Helicobacter pylori*: a hexameric chloride channel. *FEBS Lett*, **450** (1-2), 101-4.
- JACKSON, S. P. and BARTEK, J.** (2010) The DNA-damage response in human biology and disease. *Nature*, **461** (7267), 1071-1078.
- JAISWAL, A. S. and NARAYAN, S.** (2008) A novel function of adenomatous polyposis coli (APC) in regulating DNA repair. *Cancer Lett*, **271** (2), 272-80.
- JAYASEKARA, H., et al.** (2017) Risk factors for metachronous colorectal cancer or polyp: A systematic review and meta-analysis. *J Gastroenterol Hepatol*, **32** (2), 301-326.
- JENKINS, R. T., et al.** (1988) Small bowel and colonic permeability to 51Cr-EDTA in patients with active inflammatory bowel disease. *Clin Invest Med*, **11** (2), 151-5.
- JUBELIN, G., et al.** (2009) Cycle inhibiting factors (CIFs) are a growing family of functional cyclomodulins present in invertebrate and mammal bacterial pathogens. *PLoS One*, **4** (3), e4855.
- JUBELIN, G., et al.** (2010) Pathogenic bacteria target NEDD8-conjugated cullins to hijack host-cell signaling pathways. *PLoS Pathog*, **6** (9), e1001128.
- JUNG, S. W., SUGIMOTO, M., SHIOTA, S., GRAHAM, D. Y. & YAMAOKA, Y.** (2012) The intact *dupA* cluster is a more reliable *Helicobacter pylori* virulence marker than *dupA* alone. *Infect Immun*, **80** (1), 381-7.
- KAMPHUIS, J. B. J., MERCIER-BONIN, M., EUTAMENE, H. & THEODOROU, V.** (2017) Mucus organisation is shaped by colonic content; a new view. *Sci Rep*, **7** (1), 8527.
- KANEHISA, M., GOTO, S., SATO, Y., FURUMICHI, M. & TANABE, M.** (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res*, **40** (Database issue), D109-14.
- KAPLAN, C. W., LUX, R., HAAKE, S. K. & SHI, W. Y.** (2009) The *Fusobacterium nucleatum* outer membrane protein RadD is an arginine-inhibitable adhesin required for inter-species adherence and the structured architecture of multispecies biofilm. *Molecular Microbiology*, **71** (1), 35-47.
- KASPER, D. L., et al.** (1977) Isolation and identification of encapsulated strains of *Bacteroides fragilis*. *J Infect Dis*, **136** (1), 75-81.
- KATZ, J., SAMBANDAM, V., WU, J. H., MICHALEK, S. M. & BALKOVETZ, D. F.** (2000) Characterization of *Porphyromonas gingivalis*-induced degradation of epithelial cell junctional complexes. *Infect Immun*, **68** (3), 1441-9.
- KEENAN, J. I., et al.** (2016) Screening for enterotoxigenic *Bacteroides fragilis* in stool samples. *Anaerobe*, **40**, 50-3.
- KELLY, J. K. and GABOS, S.** (1987) The pathogenesis of inflammatory polyps. *Dis Colon Rectum*, **30** (4), 251-4.
- KENNEDY, N. A., et al.** (2014) The Impact of Different DNA Extraction Kits and Laboratories upon the Assessment of Human Gut Microbiota Composition by 16S rRNA Gene Sequencing. *Plos One*, **9** (2), e88982-e88982.
- KERR, J. B.** (2010) *Functional histology*. Mosby/Elsevier.
- KIM, E. R. and CHANG, D. K.** (2014) Colorectal cancer in inflammatory bowel disease: The risk, pathogenesis, prevention and diagnosis. Baishideng Publishing Group Inc.
- KIM, M. and JHO, E.-H.** (2014) Cross-talk between Wnt/ $\beta$ -catenin and Hippo signaling pathways: a brief review. *BMB reports*, **47** (10), 540-5.
- KIMURA, M., et al.** (1999) Vacuolating cytotoxin purified from *Helicobacter pylori* causes mitochondrial damage in human gastric cells. *Microb Pathog*, **26** (1), 45-52.

- KLEPPE, K., OHTSUKA, E., KLEPPE, R., MOLINEUX, I. & KHORANA, H. G.** (1971) Studies on polynucleotides. XCVI. Repair replications of short synthetic DNA's as catalyzed by DNA polymerases. *Journal of molecular biology*, **56** (2), 341-61.
- KNOWLES, M. A. and SELBY, P.** (2005) *Introduction to the cellular and molecular biology of cancer*. New York: Oxford University Press.
- KONSAVAGE JR, W. M., KYLER, S. L., RENNOLL, S. A., JIN, G. & YOCHUM, G. S.** (2012) Wnt/beta-catenin signaling regulates Yes-associated protein (YAP) gene expression in colorectal carcinoma cells. *J Biol Chem*, **287** (15), 11730-11739.
- KOOK, J. K., et al.** (2017) Genome-Based Reclassification of *Fusobacterium nucleatum* Subspecies at the Species Level. *Curr Microbiol*, **74** (10), 1137-1147.
- KOSTIC, A. D., et al.** (2013) *Fusobacterium nucleatum* potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe*, **14** (2), 207-15.
- KOSTIC, A. D., et al.** (2012) Genomic analysis identifies association of *Fusobacterium* with colorectal carcinoma. *Genome Res*, **22** (2), 292-8.
- KOSTIC, A. D., XAVIER, R. J. & GEVERS, D.** (2014) The microbiome in inflammatory bowel disease: current status and the future ahead. *Gastroenterology*, **146** (6), 1489-99.
- KRINOS, C. M., et al.** (2001) Extensive surface diversity of a commensal microorganism by multiple DNA inversions. *Nature*, **414** (6863), 555-8.
- KUCHARZIK, T., STOLL, R., LUGERING, N. & DOMSCHKE, W.** (1995) Circulating antiinflammatory cytokine IL-10 in patients with inflammatory bowel disease (IBD). *Clin Exp Immunol*, **100** (3), 452-6.
- KUCHARZIK, T., WALSH, S. V., CHEN, J., PARKOS, C. A. & NUSRAT, A.** (2001) Neutrophil transmigration in inflammatory bowel disease is associated with differential expression of epithelial intercellular junction proteins. *American Journal of Pathology*, **159** (6), 2001-2009.
- KUHN, R., LOHLER, J., RENNICK, D., RAJEWSKY, K. & MULLER, W.** (1993) Interleukin-10-deficient mice develop chronic enterocolitis. *Cell*, **75** (2), 263-74.
- KULLBERG, M. C., et al.** (1998) *Helicobacter hepaticus* triggers colitis in specific-pathogen-free interleukin-10 (IL-10)-deficient mice through an IL-12- and gamma interferon-dependent mechanism. *Infection and Immunity*, **66** (11), 5157-5166.
- KUMAR, R., et al.** (2017) *Streptococcus gallolyticus* subsp. *gallolyticus* promotes colorectal tumor development. *PLoS Pathog*, **13** (7), e1006440.
- KUMAR, S., STECHER, G. & TAMURA, K.** (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol*, **33** (7), 1870-4.
- KUMAR, V., COTRAN, R. S., ROBBINS, S. L. & KUMAR, V.** (1997) *Basic pathology*. W.B. Saunders Co.
- KUWAHARA, T., et al.** (2004) Genomic analysis of *Bacteroides fragilis* reveals extensive DNA inversions regulating cell surface adaptation. *Proc Natl Acad Sci U S A*, **101** (41), 14919-24.
- LAMAS, B., et al.** (2016) CARD9 impacts colitis by altering gut microbiota metabolism of tryptophan into aryl hydrocarbon receptor ligands. *Nat Med*, **22** (6), 598-605.
- LARA-TEJERO, M. and GALAN, J. E.** (2000) A bacterial toxin that controls cell cycle progression as a deoxyribonuclease I-like protein. *Science*, **290** (5490), 354-357.
- LAWRENZ, M. B., WOOTEN, R. M. & NORRIS, S. J.** (2004) Effects of *vlsE* complementation on the infectivity of *Borrelia burgdorferi* lacking the linear plasmid *lp28-1*. *Infect Immun*, **72** (11), 6577-85.
- LAX, A. J.** (2005) Opinion: Bacterial toxins and cancer--a case to answer? *Nat Rev Microbiol*, **3** (4), 343-9.
- LAX, A. J. and THOMAS, W.** (2002) How bacteria could cause cancer: one step at a time. *Trends in Microbiology*, **10** (6), 293-299.
- LECH, G., SŁOTWINSKI, R., SŁODKOWSKI, M. & KRASNODEBSKI, I. W.** (2016) Colorectal cancer tumour markers and biomarkers: Recent therapeutic advances. *World J Gastroenterol*, **22** (5), 1745-55.
- LEE, S. M., et al.** (2013) Bacterial colonization factors control specificity and stability of the gut microbiota. *Nature*, **501** (7467), 426-9.
- LEE, Y., EUN, C. S. & HAN, D. S.** (2016) Clinical *Fusobacterium mortiferum* Isolates Cluster with Undifferentiated *Clostridium rectum* Species Based on 16S rRNA Gene Phylogenetic Analysis. *Ann Clin Lab Sci*, **46** (3), 279-81.
- LEUNK, R. D., JOHNSON, P. T., DAVID, B. C., KRAFT, W. G. & MORGAN, D. R.** (1988) Cytotoxic activity in broth-culture filtrates of *Campylobacter pylori*. *J Med Microbiol*, **26** (2), 93-9.
- LI, J., et al.** (2014) An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol*, **32** (8), 834-41.
- LICHTENBERGER, L. M.** (1995) The hydrophobic barrier properties of gastrointestinal mucus. *Annu Rev Physiol*, **57**, 565-83.
- LINDBERG, B., LINDQVIST, B., LONNGREN, J. & POWELL, D. A.** (1980) Structural studies of the capsular polysaccharide from *Streptococcus pneumoniae* type 1. *Carbohydr Res*, **78** (1), 111-7.



- LINNEKAMP, J. F., WANG, X., MEDEMA, J. P. & VERMEULEN, L.** (2015) Colorectal cancer heterogeneity and targeted therapy: a case for molecular disease subtypes. *Cancer Res*, **75** (2), 245-9.
- LIU, J., et al.** (2013) Association of presence/absence and on/off patterns of *Helicobacter pylori* *oipA* gene with peptic ulcer disease and gastric cancer risks: a meta-analysis. *BMC Infect Dis*, **13**, 555.
- LLOYD-PRICE, J., et al.** (2017) Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature*, **550** (7674), 61-66.
- LODISH, et al.** (2013) *Molecular cell biology*.
- LU, H., HSU, P. I., GRAHAM, D. Y. & YAMAOKA, Y.** (2005) Duodenal ulcer promoting gene of *Helicobacter pylori*. *Gastroenterology*, **128** (4), 833-48.
- LU, H., et al.** (2005) Regulation of interleukin-6 promoter activation in gastric epithelial cells infected with *Helicobacter pylori*. *Mol Biol Cell*, **16** (10), 4954-66.
- LU, R., et al.** (2014) Enteric bacterial protein AvrA promotes colonic tumorigenesis and activates colonic beta-catenin signaling pathway. *Oncogenesis*, **3**, e105.
- LYNCH, H. T. and KRUSH, A. J.** (1971) Cancer family "G" revisited: 1895-1970. *Cancer*, **27** (6), 1505-11.
- MACDONALD, B. T., TAMAI, K. & HE, X.** (2009) Wnt/beta-catenin signaling: components, mechanisms, and diseases. *Dev Cell*, **17** (1), 9-26.
- MACPHERSON, A. J. and MCCOY, K. D.** (2015) Independence Day for IgA. *Immunity*, **43** (3), 416-8.
- MAGRO, F., et al.** (2013) European consensus on the histopathology of inflammatory bowel disease. *J Crohns Colitis*, **7** (10), 827-51.
- MAIDAK, B. L., et al.** (1999) A new version of the RDP (Ribosomal Database Project). *Nucleic Acids Res*, **27** (1), 171-3.
- MANSON MCGUIRE, A., et al.** (2014) Evolution of invasion in a diverse set of *Fusobacterium* species. *MBio*, **5** (6), e01864.
- MARCHESI, J. R., et al.** (2011) Towards the human colorectal cancer microbiome. *PLoS One*, **6** (5), e20447.
- MARK WELCH, J. L., ROSSETTI, B. J., RIEKEN, C. W., DEWHIRST, F. E. & BORISY, G. G.** (2016) Biogeography of a human oral microbiome at the micron scale. *Proc Natl Acad Sci U S A*, **113** (6), E791-800.
- MARSHALL, B. J., MCGECHIE, D. B., ROGERS, P. A. & GLANCY, R. J.** (1985) Pyloric *Campylobacter* infection and gastroduodenal disease. *Med J Aust*, **142** (8), 439-44.
- MARTINS, M., et al.** (2015) *Streptococcus gallolyticus* Pil3 Pilus Is Required for Adhesion to Colonic Mucus and for Colonization of Mouse Distal Colon. *J Infect Dis*, **212** (10), 1646-55.
- MARTINS, M., et al.** (2016) The Pil3 pilus of *Streptococcus gallolyticus* binds to intestinal mucins and to fibrinogen. *Gut Microbes*, **7** (6), 526-532.
- MAZMANIAN, S. K., LIU, C. H., TZIANABOS, A. O. & KASPER, D. L.** (2005) An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell*, **122** (1), 107-18.
- MAZMANIAN, S. K., ROUND, J. L. & KASPER, D. L.** (2008) A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature*, **453** (7195), 620-5.
- MCPHERSON, M. J. and MOLLER, S. G.** (2006) *PCR: The basics*. (Basics Routledge (Firm)) Taylor & Francis Group.
- MCSWEENEY, L. A. and DREYFUS, L. A.** (2004) Nuclear localization of the *Escherichia coli* cytolethal distending toxin CdtB subunit. *Cell Microbiol*, **6** (5), 447-58.
- MERA, R., et al.** (2005) Long term follow up of patients treated for *Helicobacter pylori* infection. *Gut*, **54** (11), 1536-40.
- MICHAUD, D. S., et al.** (2013) Plasma antibodies to oral bacteria and risk of pancreatic cancer in a large European prospective cohort study. *Gut*, **62** (12), 1764-70.
- MIEHLKE, S., et al.** (2000) The *Helicobacter pylori* *vacA* s1, m1 genotype and *cagA* is associated with gastric carcinoma in Germany. *Int J Cancer*, **87** (3), 322-7.
- MIMA, K., et al.** (2016) *Fusobacterium nucleatum* in Colorectal Carcinoma Tissue According to Tumor Location. *Clin Transl Gastroenterol*, **7** (11), e200.
- MIMORI-KIYOSUE, Y., SHIINA, N. & TSUKITA, S.** (2000) Adenomatous polyposis coli (APC) protein moves along microtubules and concentrates at their growing ends in epithelial cells. *Journal of Cell Biology*, **148** (3), 505-517.
- MIRAGLIA, A. G., et al.** (2007) Cytotoxic necrotizing factor 1 prevents apoptosis via the Akt/IkappaB kinase pathway: role of nuclear factor-kappaB and Bcl-2. *Mol Biol Cell*, **18** (7), 2735-44.
- MITEVA, L. D., STANILOV, N. S., DELIYSKY, T. S. & STANILOVA, S. A.** (2014) Significance of -1082A/G polymorphism of IL10 gene for progression of colorectal cancer and IL-10 expression. *Tumour Biol*, **35** (12), 12655-64.
- MONCRIEF, J. S., DUNCAN, A. J., WRIGHT, R. L., BARROSO, L. A. & WILKINS, T. D.** (1998) Molecular characterization of the fragilysin pathogenicity islet of enterotoxigenic *Bacteroides fragilis*. *Infect Immun*, **66** (4), 1735-9.

- MORIKAWA, H., et al.** (2010) The bacterial effector Cif interferes with SCF ubiquitin ligase function by inhibiting deneddylation of Cullin1. *Biochem Biophys Res Commun*, **401** (2), 268-74.
- MUISE, A. M., et al.** (2009) Polymorphisms in E-cadherin (CDH1) result in a mis-localised cytoplasmic protein that is associated with Crohn's disease. *Gut*, **58** (8), 1121-1127.
- MURATA-KAMIYA, N., et al.** (2007) *Helicobacter pylori* CagA interacts with E-cadherin and deregulates the beta-catenin signal that promotes intestinal transdifferentiation in gastric epithelial cells. *Oncogene*, **26** (32), 4617-26.
- NAGASE, H. and NAKAMURA, Y.** (1993) Mutations of the APC (adenomatous polyposis coli) gene. *Hum Mutat*, **2** (6), 425-34.
- NAKAYAMA, M., et al.** (2009) *Helicobacter pylori* VacA-induced inhibition of GSK3 through the PI3K/Akt signaling pathway. *J Biol Chem*, **284** (3), 1612-9.
- NAKAYAMA, M., et al.** (2004) *Helicobacter pylori* VacA activates the p38/activating transcription factor 2-mediated signal pathway in AZ-521 cells. *J Biol Chem*, **279** (8), 7024-8.
- NAZEMALHOSSEINI MOJARAD, E., KUPPEN, P. J., AGHDAEI, H. A. & ZALI, M. R.** (2013) The CpG island methylator phenotype (CIMP) in colorectal cancer. *Gastroenterol Hepatol Bed Bench*, **6** (3), 120-8.
- NEFF, C. P., et al.** (2016) Diverse Intestinal Bacteria Contain Putative Zwitterionic Capsular Polysaccharides with Anti-inflammatory Properties. *Cell Host Microbe*, **20** (4), 535-547.
- NGUYEN, T. L., VIEIRA-SILVA, S., LISTON, A. & RAES, J.** (2015) How informative is the mouse for human gut microbiota research? *Dis Model Mech*, **8** (1), 1-16.
- NIKITINA, A. S., et al.** (2015) Complete Genome Sequence of an Enterotoxigenic *Bacteroides fragilis* Clinical Isolate. *Genome Announc*, **3** (3), e00450-15.
- NOUGAYREDE, J. P., et al.** (2006) *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. *Science*, **313** (5788), 848-51.
- OBISO, R. J., JR., LYERLY, D. M., VAN TASSELL, R. L. & WILKINS, T. D.** (1995) Proteolytic activity of the *Bacteroides fragilis* enterotoxin causes fluid secretion and intestinal damage in vivo. *Infect Immun*, **63** (10), 3820-6.
- O'CONNOR, P. M., et al.** (2011) *Helicobacter pylori* activates calpain via toll-like receptor 2 to disrupt adherens junctions in human gastric epithelial cells. *Infect Immun*, **79** (10), 3887-94.
- ODAMAKI, T., et al.** (2008) Distribution of different species of the *Bacteroides fragilis* group in individuals with Japanese cedar pollinosis. *Appl Environ Microbiol*, **74** (21), 6814-7.
- ODENBREIT, S., et al.** (2000) Translocation of *Helicobacter pylori* CagA into gastric epithelial cells by type IV secretion. *Science*, **287** (5457), 1497-1500.
- ODENWALD, M. A. and TURNER, J. R.** (2017) The intestinal epithelial barrier: a therapeutic target? *Nat Rev Gastroenterol Hepatol*, **14** (1), 9-21.
- O'HARA, R. J., et al.** (1998) Advanced colorectal cancer is associated with impaired interleukin 12 and enhanced interleukin 10 production. *Clin Cancer Res*, **4** (8), 1943-8.
- OHKUSA, T., et al.** (2003) Induction of experimental ulcerative colitis by *Fusobacterium varium* isolated from colonic mucosa of patients with ulcerative colitis. *Gut*, **52** (1), 79-83.
- PAPINI, E., et al.** (1994) Cellular vacuoles induced by *Helicobacter pylori* originate from late endosomal compartments. *Proc Natl Acad Sci U S A*, **91** (21), 9720-4.
- PAREJA, M. E. and COLOMBO, M. I.** (2013) Autophagic clearance of bacterial pathogens: molecular recognition of intracellular microorganisms. *Front Cell Infect Microbiol*, **3**, 54.
- PARKES, M., et al.** (2007) Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat Genet*, **39** (7), 830-2.
- PARSONNET, J., et al.** (1991) *Helicobacter pylori* infection and the risk of gastric carcinoma. *N Engl J Med*, **325** (16), 1127-31.
- PASCAL, V., et al.** (2017) A microbial signature for Crohn's disease. *Gut*, **66** (5), 813-822.
- PATRICK, S. and BLAKELY, G. W.** (2012) Crossing the eukaryote-prokaryote divide: A ubiquitin homolog in the human commensal bacterium *Bacteroides fragilis*. *Mob Genet Elements*, **2** (3), 149-151.
- PATRICK, S., et al.** (2010) Twenty-eight divergent polysaccharide loci specifying within- and amongst-strain capsule diversity in three strains of *Bacteroides fragilis*. *Microbiology*, **156** (Pt 11), 3255-69.
- PATRICK, S., et al.** (2011) A unique homologue of the eukaryotic protein-modifier ubiquitin present in the bacterium *Bacteroides fragilis*, a predominant resident of the human gastrointestinal tract. *Microbiology*, **157** (Pt 11), 3071-8.
- PATRICK, S., et al.** (2003) Multiple inverted DNA repeats of *Bacteroides fragilis* that control polysaccharide antigenic variation are similar to the hin region inverted repeats of *Salmonella typhimurium*. *Microbiology*, **149** (Pt 4), 915-24.
- PENG, L., LI, Z. R., GREEN, R. S., HOLZMAN, I. R. & LIN, J.** (2009) Butyrate enhances the intestinal barrier by facilitating tight junction assembly via activation of AMP-activated protein kinase in Caco-2 cell monolayers. *J Nutr*, **139** (9), 1619-25.

- PETERSON, L. W. and ARTIS, D.** (2014) Intestinal epithelial cells: regulators of barrier function and immune homeostasis. *Nat Rev Immunol*, **14** (3), 141-53.
- PETERSSON, J., et al.** (2011) Importance and regulation of the colonic mucus barrier in a mouse model of colitis. *Am J Physiol Gastrointest Liver Physiol*, **300** (2), G327-33.
- PICARD, B., et al.** (1999) The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. *Infect Immun*, **67** (2), 546-53.
- PIERCE, J. V. and BERNSTEIN, H. D.** (2016) Genomic Diversity of Enterotoxigenic Strains of *Bacteroides fragilis*. *PLoS One*, **11** (6), e0158171.
- PIGNEUR, B., et al.** (2013) Phenotypic characterization of very early-onset IBD due to mutations in the IL10, IL10 receptor alpha or beta gene: a survey of the Genius Working Group. *Inflamm Bowel Dis*, **19** (13), 2820-8.
- PIKOR, L., THU, K., VUCIC, E. & LAM, W.** (2013) The detection and implication of genome instability in cancer. *Cancer Metastasis Rev*, **32** (3-4), 341-52.
- POON, R. Y., CHAU, M. S., YAMASHITA, K. & HUNTER, T.** (1997) The role of Cdc2 feedback loop control in the DNA damage checkpoint in mammalian cells. *Cancer Res*, **57** (22), 5168-78.
- PRINDIVILLE, T. P., et al.** (2000) *Bacteroides fragilis* enterotoxin gene sequences in patients with inflammatory bowel disease. *Emerg Infect Dis*, **6** (2), 171-4.
- PURCELL, R. V., et al.** (2017) Colonization with enterotoxigenic *Bacteroides fragilis* is associated with early-stage colorectal neoplasia. *PLoS One*, **12** (2), e0171602.
- PURCELL, R. V., PEARSON, J., FRIZELLE, F. A. & KEENAN, J. I.** (2016) Comparison of standard, quantitative and digital PCR in the detection of enterotoxigenic *Bacteroides fragilis*. *Sci Rep*, **6** (1), 34554.
- QIN, J., et al.** (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, **464** (7285), 59-65.
- QUAST, C., et al.** (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*, **41** (Database issue), D590-6.
- RAISCH, J., et al.** (2014) Colon cancer-associated B2 *Escherichia coli* colonize gut mucosa and promote cell proliferation. *World J Gastroenterol*, **20** (21), 6560-72.
- RAMPIAS, T., et al.** (2010) Activation of Wnt signaling pathway by human papillomavirus E6 and E7 oncogenes in HPV16-positive oropharyngeal squamous carcinoma cells. *Mol Cancer Res*, **8** (3), 433-43.
- RANDI, G., FRANCESCHI, S. & LA VECCHIA, C.** (2006) Gallbladder cancer worldwide: geographical distribution and risk factors. *Int J Cancer*, **118** (7), 1591-602.
- RAO, C. V. and YAMADA, H. Y.** (2013) Genomic instability and colon carcinogenesis: from the perspective of genes. *Front Oncol*, **3**, 130.
- REMACLE, A. G., SHIRYAEV, S. A. & STRONGIN, A. Y.** (2014) Distinct interactions with cellular E-cadherin of the two virulent metalloproteinases encoded by a *Bacteroides fragilis* pathogenicity island. *PLoS One*, **9** (11), e113896.
- RHEE, K. J., et al.** (2009) Induction of persistent colitis by a human commensal, enterotoxigenic *Bacteroides fragilis*, in wild-type C57BL/6 mice. *Infect Immun*, **77** (4), 1708-18.
- RICHARD, M. L., et al.** (2017) Mucosa-associated microbiota dysbiosis in colitis associated cancer. *Gut Microbes*. Taylor & Francis.
- ROSTY, C. and BETTINGTON, M.** (2014) Serrated colorectal polyps and polyposis. *Diagnostic Histopathology*, **20** (1), 30-37.
- ROUND, J. L., et al.** (2011) The Toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota. *Science*, **332** (6032), 974-7.
- ROWAN, A. J., et al.** (2000) APC mutations in sporadic colorectal tumors: A mutational "hotspot" and interdependence of the "two hits". *Proc Natl Acad Sci U S A*, **97** (7), 3352-7.
- RUBINSTEIN, M. R., et al.** (2013) *Fusobacterium nucleatum* promotes colorectal carcinogenesis by modulating E-cadherin/beta-catenin signaling via its FadA adhesin. *Cell Host Microbe*, **14** (2), 195-206.
- RUDDON, R. W.** (2007) *Cancer biology*. Oxford ; New York: Oxford University Press.
- RUDKIN, J. K., MCLOUGHLIN, R. M., PRESTON, A. & MASSEY, R. C.** (2017) Bacterial toxins: Offensive, defensive, or something else altogether? *PLoS Pathog*, **13** (9), e1006452.
- RUTTER, M., et al.** (2004) Severity of inflammation is a risk factor for colorectal neoplasia in ulcerative colitis. *Gastroenterology*, **126** (2), 451-459.
- SABINO, J., et al.** (2016) Primary sclerosing cholangitis is characterised by intestinal dysbiosis independent from IBD. *Gut*, **65** (10), 1681-9.
- SAIDI, R. F., JAEGER, K., MONTROSE, M. H., WU, S. G. & SEARS, C. L.** (1997) *Bacteroides fragilis* toxin rearranges the actin cytoskeleton of HT29/C1 cells without direct proteolysis of actin or decrease in F-actin content. *Cell Motility and the Cytoskeleton*, **37** (2), 159-165.
- SAIKI, R. K., et al.** (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239** (4839), 487-91.

- SAITOU, N. and NEI, M.** (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*, **4** (4), 406-25.
- SAMBA-LOUAKA, A., et al.** (2008) Bacterial cyclomodulin Cif blocks the host cell cycle by stabilizing the cyclin-dependent kinase inhibitors p21 and p27. *Cell Microbiol*, **10** (12), 2496-508.
- SAMBA-LOUAKA, A., et al.** (2014) *Listeria monocytogenes* dampens the DNA damage response. *PLoS Pathog*, **10** (10), e1004470.
- SANTAOLALLA, R., et al.** (2013) TLR4 activates the beta-catenin pathway to cause intestinal neoplasia. *PLoS One*, **8** (5), e63298.
- SATYANARAYANA, A. and KALDIS, P.** (2009) Mammalian cell-cycle regulation: several Cdks, numerous cyclins and diverse compensatory mechanisms. *Oncogene*, **28** (33), 2925-39.
- SCALDAFERRI, F. and FIOCCHI, C.** (2007) Inflammatory bowel disease: progress and current concepts of etiopathogenesis. *J Dig Dis*, **8** (4), 171-8.
- SCANU, T., et al.** (2015) *Salmonella* Manipulation of Host Signaling Pathways Provokes Cellular Transformation Associated with Gallbladder Carcinoma. *Cell Host Microbe*, **17** (6), 763-74.
- SCHLEGEL, L., GRIMONT, F., AGERON, E., GRIMONT, P. A. & BOUVET, A.** (2003) Reappraisal of the taxonomy of the *Streptococcus bovis*/*Streptococcus equinus* complex and related species: description of *Streptococcus gallolyticus* subsp. *gallolyticus* subsp. nov., *S. gallolyticus* subsp. *macedonicus* subsp. nov. and *S. gallolyticus* subsp. *pasteurianus* subsp. nov. *Int J Syst Evol Microbiol*, **53** (Pt 3), 631-45.
- SCHMIDT, G., SELZER, J., LERM, M. & AKTORIES, K.** (1998) The Rho-deamidating cytotoxic necrotizing factor 1 from *Escherichia coli* possesses transglutaminase activity. Cysteine 866 and histidine 881 are essential for enzyme activity. *J Biol Chem*, **273** (22), 13669-74.
- SCHNEIDER, M. R., et al.** (2010) A key role for E-cadherin in intestinal homeostasis and Paneth cell maturation. *PLoS One*, **5** (12), e14325.
- SCHRADER, C., SCHIELKE, A., ELLERBROEK, L. & JOHNE, R.** (2012) PCR inhibitors - occurrence, properties and removal. *J Appl Microbiol*, **113** (5), 1014-26.
- SEARS, C. L. and PARDOLL, D. M.** (2011) Perspective: alpha-bugs, their microbial partners, and the link to colon cancer. *J Infect Dis*, **203** (3), 306-11.
- SEGAL, E., HAGBLOM, P., SEIFERT, H. S. & SO, M.** (1986) Antigenic variation of gonococcal pilus involves assembly of separated silent gene segments. *Proc Natl Acad Sci U S A*, **83** (7), 2177-81.
- SELANDER, R. K., et al.** (1986) Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl Environ Microbiol*, **51** (5), 873-84.
- SHIMPOH, T., et al.** (2017) Prevalence of pks-positive *Escherichia coli* in Japanese patients with or without colorectal cancer. *Gut Pathog*, **9** (1), 35.
- SHIOTA, S., MATSUNARI, O., WATADA, M., HANADA, K. & YAMAOKA, Y.** (2010) Systematic review and meta-analysis: the relationship between the *Helicobacter pylori* *dupA* gene and clinical outcomes. *Gut Pathog*, **2** (1), 13.
- SHIRYAEV, S. A., et al.** (2014) Structural and functional diversity of metalloproteinases encoded by the *Bacteroides fragilis* pathogenicity island. *FEBS J*, **281** (11), 2487-502.
- SHOUVAL, D. S., et al.** (2014) Interleukin 10 receptor signaling: master regulator of intestinal mucosal homeostasis in mice and humans. *Adv Immunol*, **122**, 177-210.
- SIEBER, O. M., et al.** (2002) Analysis of chromosomal instability in human colorectal adenomas with two mutational hits at APC. *Proc Natl Acad Sci U S A*, **99** (26), 16910-5.
- SIEGL, C., PRUSTY, B. K., KARUNAKARAN, K., WISCHHUSEN, J. & RUDEL, T.** (2014) Tumor suppressor p53 alters host cell metabolism to limit *Chlamydia trachomatis* infection. *Cell Rep*, **9** (3), 918-29.
- SMITH, P. M., et al.** (2013) The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis. *Science*, **341** (6145), 569-73.
- SOKOL, H., et al.** (2008) *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn's disease patients. *Proc Natl Acad Sci U S A*, **105** (43), 16731-6.
- SOKURENKO, E. V., GOMULKIEWICZ, R. & DYKHUIZEN, D. E.** (2006) Source-sink dynamics of virulence evolution. *Nat Rev Microbiol*, **4** (7), 548-55.
- SONNENSCHN, C. and SOTO, A. M.** (2013) The aging of the 2000 and 2011 Hallmarks of Cancer reviews: a critique. *J Biosci*, **38** (3), 651-63.
- SPENCER, S. D., et al.** (1998) The orphan receptor CRF2-4 is an essential subunit of the interleukin 10 receptor. *Journal of Experimental Medicine*, **187** (4), 571-578.
- STANILOV, N., et al.** (2010) [Role of IL-12P40 and IL-10 in progression of colorectal cancer]. *Khirurgiia (Sofia)*, (4-5), 26-9
- STEIN, M., et al.** (2002) c-Src/Lyn kinases activate *Helicobacter pylori* CagA through tyrosine phosphorylation of the EPIYA motifs. *Mol Microbiol*, **43** (4), 971-80.

- STRAUSS, J., et al.** (2011) Invasive potential of gut mucosa-derived *Fusobacterium nucleatum* positively correlates with IBD status of the host. *Inflamm Bowel Dis*, **17** (9), 1971-8.
- STRAUSS, J., WHITE, A., AMBROSE, C., MCDONALD, J. & ALLEN-VERCOE, E.** (2008) Phenotypic and genotypic analyses of clinical *Fusobacterium nucleatum* and *Fusobacterium periodonticum* isolates from the human gut. *Anaerobe*, **14** (6), 301-9.
- SU, L. K., et al.** (1992) Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene. *Science*, **256** (5057), 668-70.
- SUGANUMA, M., et al.** (2008) TNF-alpha-inducing protein, a carcinogenic factor secreted from *H. pylori*, enters gastric cancer cells. *Int J Cancer*, **123** (1), 117-22.
- SURANA, N. K. and KASPER, D. L.** (2012) The yin yang of bacterial polysaccharides: lessons learned from *B. fragilis* PSA. *Immunol Rev*, **245** (1), 13-26.
- SZKARADKIEWICZ, A., et al.** (2009) Proinflammatory cytokines and IL-10 in inflammatory bowel disease and colorectal cancer patients. *Arch Immunol Ther Exp (Warsz)*, **57** (4), 291-4.
- TAGORE, A., et al.** (1999) Interleukin-10 (IL-10) genotypes in inflammatory bowel disease. *Tissue Antigens*, **54** (4), 386-90.
- TAHARA, T., et al.** (2015) *Fusobacterium* detected in colonic biopsy and clinicopathological features of ulcerative colitis in Japan. *Dig Dis Sci*, **60** (1), 205-10.
- TAHARA, T., et al.** (2014) *Fusobacterium* in colonic flora and molecular features of colorectal carcinoma. *Cancer Res*, **74** (5), 1311-8.
- TAMURA, K., NEI, M. & KUMAR, S.** (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci U S A*, **101** (30), 11030-5.
- TANAKA, H., et al.** (2015) Intestinal deletion of Claudin-7 enhances paracellular organic solute flux and initiates colonic inflammation in mice. *Gut*, **64** (10), 1529-38.
- TANNOCK, G. W., et al.** (2012) Comprehensive analysis of the bacterial content of stool from patients with chronic pouchitis, normal pouches, or familial adenomatous polyposis pouches. *Inflamm Bowel Dis*, **18** (5), 925-34.
- TEGTMEYER, N., et al.** (2017) *Helicobacter pylori* Employs a Unique Basolateral Type IV Secretion Mechanism for CagA Delivery. *Cell Host Microbe*, **22** (4), 552-560 e5.
- TETTELIN, H., et al.** (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc Natl Acad Sci U S A*, **102** (39), 13950-5.
- TJALSMA, H., BOLEIJ, A., MARCHESI, J. R. & DUTILH, B. E.** (2012) A bacterial driver-passenger model for colorectal cancer: beyond the usual suspects. *Nat Rev Microbiol*, **10** (8), 575-82.
- TOMBOLA, F., et al.** (1999) *Helicobacter pylori* vacuolating toxin forms anion-selective channels in planar lipid bilayers: possible implications for the mechanism of cellular vacuolation. *Biophys J*, **76** (3), 1401-9.
- TOMKOVICH, S., et al.** (2017) Locoregional Effects of Microbiota in a Preclinical Model of Colon Carcinogenesis. *Cancer Res*, **77** (10), 2620-2632.
- TOPRAK, N. U., et al.** (2006) A possible role of *Bacteroides fragilis* enterotoxin in the aetiology of colorectal cancer. *Clin Microbiol Infect*, **12** (8), 782-6.
- TRAVASSOS, L. H., et al.** (2010) Nod1 and Nod2 direct autophagy by recruiting ATG16L1 to the plasma membrane at the site of bacterial entry. *Nat Immunol*, **11** (1), 55-62.
- TROY, E. B., CAREY, V. J., KASPER, D. L. & COMSTOCK, L. E.** (2010) Orientations of the *Bacteroides fragilis* capsular polysaccharide biosynthesis locus promoters during symbiosis and infection. *J Bacteriol*, **192** (21), 5832-6.
- UEMURA, N., et al.** (2001) *Helicobacter pylori* infection and the development of gastric cancer. *N Engl J Med*, **345** (11), 784-9.
- UNTERHOLZNER, S. J., POPPENBERGER, B. & ROZHON, W.** (2013) Toxin-antitoxin systems: Biology, identification, and application. *Mob Genet Elements*, **3** (5), e26219.
- VAISHNAVA, S., et al.** (2011) The antibacterial lectin RegIIIgamma promotes the spatial segregation of microbiota and host in the intestine. *Science*, **334** (6053), 255-8.
- VAN DER SLUIS, M., et al.** (2006) Muc2-deficient mice spontaneously develop colitis, indicating that MUC2 is critical for colonic protection. *Gastroenterology*, **131** (1), 117-29.
- VAN DER WAAIJ, L. A., et al.** (2005) Bacterial population analysis of human colon and terminal ileum biopsies with 16S rRNA-based fluorescent probes: commensal bacteria live in suspension and have no direct contact with epithelial cells. *Inflamm Bowel Dis*, **11** (10), 865-71.
- VAN DER WOUDE, M. W. and BÄUMLER, A. J.** (2004) Phase and antigenic variation in bacteria. American Society for Microbiology (ASM).
- VAN HAM, S. M., VAN ALPHEN, L., MOOI, F. R. & VAN PUTTEN, J. P. M.** (1993) Phase variation of *H. influenzae* fimbriae: Transcriptional control of two divergent genes through a variable combined promoter region. *Cell*, **73** (6), 1187-1196.

- VAN HEEL, D. A., et al.** (2003) The IBD6 Crohn's disease locus demonstrates complex interactions with CARD15 and IBD5 disease-associated variants. *Hum Mol Genet*, **12** (20), 2569-75.
- VASU, K. and NAGARAJA, V.** (2013) Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol Mol Biol Rev*, **77** (1), 53-72.
- VILJOEN, K. S., DAKSHINAMURTHY, A., GOLDBERG, P. & BLACKBURN, J. M.** (2015) Quantitative profiling of colorectal cancer-associated bacteria reveals associations between *Fusobacterium* spp., enterotoxigenic *Bacteroides fragilis* (ETBF) and clinicopathological features of colorectal cancer. *PLoS One*, **10** (3), e0119462.
- VIZCAINO, M. I. and CRAWFORD, J. M.** (2015) The colibactin warhead crosslinks DNA. *Nat Chem*, **7** (5), 411-7.
- VOGELSTEIN, B., et al.** (1988) Genetic alterations during colorectal-tumor development. *N Engl J Med*, **319** (9), 525-32.
- VOGELSTEIN, B., et al.** (2013) Cancer genome landscapes. *Science*, **339** (6127), 1546-58.
- WANG, A. H., et al.** (2011) The effect of IL-10 genetic variation and interleukin 10 serum levels on Crohn's disease susceptibility in a New Zealand population. *Hum Immunol*, **72** (5), 431-5.
- WANG, M. Y., et al.** (2015) Intact long-type DupA protein in *Helicobacter pylori* is an ATPase involved in multifunctional biological activities. *Microb Pathog*, **81**, 53-9.
- WANG, X., et al.** (2012) 4-hydroxy-2-nonenal mediates genotoxicity and bystander effects caused by *Enterococcus faecalis*-infected macrophages. *Gastroenterology*, **142** (3), 543-551 e7.
- WANG, Y., et al.** (2017) Safety Evaluation of a Novel Strain of *Bacteroides fragilis*. *Front Microbiol*, **8**, 435.
- WANG, Y., KALKA-MOLL, W. M., ROEHRL, M. H. & KASPER, D. L.** (2000) Structural basis of the abscess-modulating polysaccharide A2 from *Bacteroides fragilis*. *Proc Natl Acad Sci U S A*, **97** (25), 13478-83.
- WARBURG, O.** (1956) On the Origin of Cancer Cells. *Science*, **123** (3191), 309-314.
- WARD, P. A. and LENTSCH, A. B.** (1999) The acute inflammatory response and its regulation. *Arch Surg*, **134** (6), 666-9.
- WARREN, R. L., et al.** (2013) Co-occurrence of anaerobic bacteria in colorectal carcinomas. *Microbiome*, **1** (1), 16.
- WATSON, A. J. and HUGHES, K. R.** (2012) TNF-alpha-induced intestinal epithelial cell shedding: implications for intestinal barrier function. *Ann N Y Acad Sci*, **1258** (1), 1-8.
- WEBER-NORDT, R. M., et al.** (1996) Stat3 Recruitment by Two Distinct Ligand-induced, Tyrosine-phosphorylated Docking Sites in the Interleukin-10 Receptor Intracellular Domain. *Journal of Biological Chemistry*, **271** (44), 27954-27961.
- WEINBERG, R. A.** (2014) *The biology of cancer*. New York: Garland Science, Taylor & Francis Group.
- WEISER, J. N., AUSTRIAN, R., SREENIVASAN, P. K. & MASURE, H. R.** (1994) Phase variation in pneumococcal opacity: relationship between colonial morphology and nasopharyngeal colonization. *Infect Immun*, **62** (6), 2582-9.
- WEXLER, H. M.** (2007) *Bacteroides*: the good, the bad, and the nitty-gritty. *Clin Microbiol Rev*, **20** (4), 593-621.
- WICK, E. C., et al.** (2014) Stat3 activation in murine colitis induced by enterotoxigenic *Bacteroides fragilis*. *Inflamm Bowel Dis*, **20** (5), 821-34.
- WIEDBRAUK, D. L., WERNER, J. C. & DREVON, A. M.** (1995) Inhibition of PCR by aqueous and vitreous fluids. *J Clin Microbiol*, **33** (10), 2643-6.
- WINAWER, S. J., et al.** (1993) Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. *N Engl J Med*, **329** (27), 1977-81.
- WLODARSKA, M., et al.** (2017) Indoleacrylic Acid Produced by Commensal *Peptostreptococcus* Species Suppresses Inflammation. *Cell Host Microbe*, **22** (1), 25-37 e6.
- WOESE, C. R., KANDLER, O. & WHEELIS, M. L.** (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A*, **87** (12), 4576-9.
- WONG, B. C., et al.** (2004) *Helicobacter pylori* eradication to prevent gastric cancer in a high-risk region of China: a randomized controlled trial. *JAMA*, **291** (2), 187-94.
- WU, S., LIM, K. C., HUANG, J., SAIDI, R. F. & SEARS, C. L.** (1998) *Bacteroides fragilis* enterotoxin cleaves the zonula adherens protein, E-cadherin. *Proc Natl Acad Sci U S A*, **95** (25), 14979-84.
- WU, S., MORIN, P. J., MAOUYO, D. & SEARS, C. L.** (2003) *Bacteroides fragilis* enterotoxin induces c-Myc expression and cellular proliferation. *Gastroenterology*, **124** (2), 392-400.
- WU, S., et al.** (2009) A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. *Nat Med*, **15** (9), 1016-22.
- WU, S., RHEE, K. J., ZHANG, M., FRANCO, A. & SEARS, C. L.** (2007) *Bacteroides fragilis* toxin stimulates intestinal epithelial cell shedding and gamma-secretase-dependent E-cadherin cleavage. *J Cell Sci*, **120** (Pt 11), 1944-52.

- WU, S., et al.** (2006) The *Bacteroides fragilis* toxin binds to a specific intestinal epithelial cell receptor. *Infect Immun*, **74** (9), 5382-90.
- WU, T. C. and PARK, J. T.** (1971) Chemical characterization of a new surface antigenic polysaccharide from a mutant of *Staphylococcus aureus*. *J Bacteriol*, **108** (2), 874-84.
- XU, M., et al.** (2007) FadA from *Fusobacterium nucleatum* utilizes both secreted and nonsecreted forms for functional oligomerization for attachment and invasion of host cells. *J Biol Chem*, **282** (34), 25000-9.
- YAMADA, S., POKUTTA, S., DREES, F., WEIS, W. I. & NELSON, W. J.** (2005) Deconstructing the cadherin-catenin-actin complex. *Cell*, **123** (5), 889-901.
- YAMAOKA, Y., KWON, D. H. & GRAHAM, D. Y.** (2000) A M(r) 34,000 proinflammatory outer membrane protein (oipA) of *Helicobacter pylori*. *Proc Natl Acad Sci U S A*, **97** (13), 7533-8.
- YE, X., et al.** (2017) *Fusobacterium Nucleatum* Subspecies *Animalis* Influences Proinflammatory Cytokine Expression and Monocyte Activation in Human Colorectal Tumors. *Cancer Prev Res (Phila)*, **10** (7), 398-409.
- YIM, S., et al.** (2013) Enterotoxigenic *Bacteroides fragilis* causes lethal colitis in Mongolian gerbils. *Anaerobe*, **21**, 64-6.
- YU, J., et al.** (2016) Invasive *Fusobacterium nucleatum* may play a role in the carcinogenesis of proximal colon cancer through the serrated neoplasia pathway. *Int J Cancer*, **139** (6), 1318-26.
- YU, T., et al.** (2017) *Fusobacterium nucleatum* Promotes Chemoresistance to Colorectal Cancer by Modulating Autophagy. *Cell*, **170** (3), 548-563 e16.
- ZAKHARZHEVSKAYA, N. B., et al.** (2017) Interaction of *Bacteroides fragilis* Toxin with Outer Membrane Vesicles Reveals New Mechanism of Its Secretion and Delivery. *Front Cell Infect Microbiol*, **7**, 2.
- ZAMANI, S., et al.** (2017) Detection of enterotoxigenic *Bacteroides fragilis* in patients with ulcerative colitis. *Gut Pathog*, **9** (1), 53.
- ZELLA, G. C., et al.** (2011) Distinct microbiome in pouchitis compared to healthy pouches in ulcerative colitis and familial adenomatous polyposis. *Inflamm Bowel Dis*, **17** (5), 1092-100.
- ZENG, X. T., et al.** (2016) Periodontal Disease and Incident Lung Cancer Risk: A Meta-Analysis of Cohort Studies. *J Periodontol*, **87** (10), 1158-64.
- ZHANG, B., et al.** (2014) Proteogenomic characterization of human colon and rectal cancer. *Nature*, **513** (7518), 382-7.
- ZHANG, J. R., HARDHAM, J. M., BARBOUR, A. G. & NORRIS, S. J.** (1997) Antigenic variation in Lyme disease *borreliae* by promiscuous recombination of VMP-like sequence cassettes. *Cell*, **89** (2), 275-85.
- ZHOU, Y., et al.** (2015) Noncanonical activation of beta-catenin by *Porphyromonas gingivalis*. *Infect Immun*, **83** (8), 3195-203.
- ZUMBRUNN, J., KINOSHITA, K., HYMAN, A. A. & NÄTHKE, I. S.** (2001) Binding of the adenomatous polyposis coli protein to microtubules increases microtubule stability and is regulated by GSK3 $\beta$  phosphorylation. *Current Biology*, **11** (1), 44-49.

APPENDIX A

Centre Number:

Study Number:

Patient Identification Number for this trial:

---

CONSENT FORM

---

Title of Project: **The Role of Bacterial Infections in Cancer Development**

Name of Researcher: **Dr. Jeremy Sanderson (Gastroenterology)**

**Professor Alistair Lax (Microbiology)**

Please initial all boxes

1. I confirm that I have read and understand the information sheet dated 23<sup>rd</sup> August 2016 (version 1) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.

☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my medical care or legal rights being affected.

☐

3. I understand that relevant sections of my medical notes and data collected during the study may be looked at by individuals from the Department of Gastroenterology where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records.

☐

4. I understand that samples of blood and colonic tissue and fluid donated by me will either be analysed immediately or stored until no later than September 2024 in case we need to undertake further analysis as a result of our research.

☐

5. I agree to my donated samples being used to analyse my DNA.

NO

☐

YES

☐

6. I understand that samples donated by me may be analysed anonymously in laboratories in France, Sweden, Denmark and Italy as well in the UK and stored for this research and destroyed no later than September 2024.

☐

7. I agree to take part in the above study.

☐

\_\_\_\_\_  
Name of Participant

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Name of Person  
taking consent.

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature



## APPENDIX B

Department of Gastroenterology  
1<sup>st</sup> Floor, College House, South Wing  
**St Thomas' Hospital**  
Westminster Bridge Road  
London SE1 7EH

# Patient Information Sheet

## The Role of Bacterial Infections in Cancer Development

### Part 1

We would like to invite you to take part in our research study. Before you decide, we would like you to understand why the research is being done and what it would involve for you. We will be happy to answer any questions you may have before you decide to take part. You may also wish to discuss the study with friends and family. Participation is entirely voluntary and you may withdraw from the study at any time without giving a reason

#### **Am I eligible to take part?**

We are looking for adults, aged 18 to 80 years, who are undergoing routine colonoscopy to investigate symptoms of altered bowel habit, abdominal pain or rectal bleeding or as a follow up examination for colonic polyps, bowel cancer or inflammatory bowel disease and therefore are recruiting a range of patients with a range of conditions. We are asking about 300 people to take part in the study. Participating in the study will not alter your routine care. Your eligibility to take part in the study will be confirmed on the day of your procedure by the endoscopist (doctor).

#### **What is the study for?**

We want to find out whether certain toxin-producing bacteria are involved in the development of colorectal (bowel) cancer

#### **The main objective of the study is to:**

Investigate whether certain types of bacteria that produce toxins (for example, *Bacteroides* and *E. coli*) are present in the intestine in individuals found to have bowel cancer or pre-cancerous polyps or in people with longstanding inflammation of the bowel (inflammatory bowel disease).

#### **The secondary objectives of the study are to:**

- Determine the presence of toxin producing bacteria in samples of the bowel lining taken during routine colonoscopy
- Determine if the presence of toxin producing bacteria is linked to the presence of polyps, bowel cancer or bowel inflammation
- Determine the presence of toxin producing bacteria in faecal contents taken from the right colon during colonoscopy
- Assess the relationship between bacterial toxins and the cells of the immune system in the gut and blood

#### **Do I have to take part?**

It is up to you if you decide to take part in the study. We will describe the study and go through this information sheet. If you agree to take part, we will then ask you to sign a consent form. You are free to withdraw at any time, without giving a reason. This will not affect the standard of care you receive.

**What will happen if I take part?**

The study involves taking some additional samples from the lining of the bowel (called biopsies) during your colonoscopy. It is routine to take numerous small biopsy samples from the lining during a colonoscopy and you do not feel this. As many as 30 of these may be taken routinely during a colonoscopy. In our study, we are asking your permission to take up to 14 biopsies additional to those for routine diagnosis from different sites in the bowel. The site will be chosen depending on the findings at your colonoscopy but will be either just from normal lining or from normal and abnormal areas such as a polyp or bowel cancer. Taking samples from abnormal areas will only be done if it has no impact on the accuracy of your medical diagnosis and treatment.

In addition, we will take a small sample of faecal liquid from the inside of the colon (aspirated up the colonoscope) which you will not feel.

We are also asking you to provide a 26 mL sample of blood (about 4 and half teaspoons) which will be taken from the cannula already placed in your arm or hand and will therefore not involve any additional needles or cause discomfort.

**What will happen to my samples?**

Your samples will be marked with your participant number, the date, sample type and sample code only, so you will not be identifiable to the researchers studying your samples. The samples of bowel lining and faecal fluid will be analysed for the presence of bacteria and bacterial toxins using molecular (DNA) and antibody-based detection methods. Bacteria will be grown from the biopsy and faecal fluid samples. The blood sample will be used to analyse the cells of the immune system and how these interact with bacteria found in the biopsy and faecal liquid samples. Molecular (genetic) analysis will be done on your own DNA (genes) to determine whether certain genes react with bacteria or bacterial toxins to make colonic polyps or cancers occur.

The samples will either be analysed immediately or be stored in a freezer in a locked laboratory on a corridor not accessible to the public. Once processed, your samples will be analysed at Kings College London (Microbiology). They may also be sent anonymously to participating research centres in Europe (Karolinska Institute, Stockholm, Sweden; University of Copenhagen, Denmark; National Institute of Health, Rome, Italy; University of Bologna, Italy; Pasteur Institute, France; Centre Hospitalier Universitaire de Toulouse, France). They will be stored until September 2024 in case we need to undertake further analysis as a result of our research into bacteria and colon cancer. Your samples will either be completely used for analysis or destroyed according to the Human Tissue Act (HTA) guidelines.

**What are the advantages/disadvantages in taking part?**

The study involves taking the type of samples done as a routine part of undergoing a colonoscopy. By taking up to 14 samples, additional to any taken for routine diagnosis, the procedure will take up to 8 minutes longer but will not be more uncomfortable. Each biopsy is painless and leads to a tiny amount of bleeding from the lining of the bowel which quickly stops. Aspirating the faecal liquid takes about half a minute. The blood sample will be taken during your recovery period so will not prolong the procedure in any way. We don't anticipate any direct benefits to participants from taking part, but it will help us understand the factors which promote the development of bowel cancer.

**What if there is a problem?**

Any complaint about the way you have been dealt with during the study or any possible harm you might suffer will be addressed. The detailed information about this is given in part 2.

**Will my taking part in this study be kept confidential?**

Yes. We will follow ethical and legal practice and all information will be handled in confidence. The details are included in part 2.

## APPENDIX C

Department of Gastroenterology  
1<sup>st</sup> Floor, College House, South Wing  
**St Thomas' Hospital**  
Westminster Bridge Road  
London SE1 7EH

# Patient Information Sheet

## Gut Bacteria in Colorectal Cancer

### Part 2

#### **What will happen if I wish to withdraw from the study?**

You are free to withdraw from the study at any time without giving a reason. If you decide to do so after the samples have been taken, we will destroy your samples and any information relating to you. If you decide to withdraw, it will not affect your normal care in any way.

#### **Who is organising and funding the study?**

The project is organised by researchers from Guy's & St. Thomas' Hospitals NHS Foundation Trust. Some funding has been awarded by the Food and Gut Research Fund, Guy's & St' Thomas' Charity and the Danish Council for Independent Research. The study has also been reviewed and given a favourable opinion by Wales REC7 (14/WA/1221) Research Ethics Committee, an independent group of people who protect the interests of research participants.

#### **What if there is a problem?**

If you have any questions/concerns about any aspect of this study, you should ask to speak to the researchers (020 7188 2497, [jeremy.sanderson@kcl.ac.uk](mailto:jeremy.sanderson@kcl.ac.uk)), who will do their best to answer your questions.

Address for correspondence:

Dept of Gastroenterology, Guy's & St. Thomas' Hospitals  
1<sup>st</sup> Floor, College House, St. Thomas' Hospital  
Westminster Bridge Road, London, SE1 7EH

If you have a complaint, you should talk to your research doctor who will do their best to answer your questions. If you remain unhappy and wish to complain formally, you can do this through the NHS complaints procedure. Details can be obtained through the Patient Liaison and Advisory Service (PALS) at Guy's & St. Thomas' Hospitals on 02071887188, address: PALS, KIC, Ground Floor, North Wing, St. Thomas' Hospital, Westminster Bridge

Road, London, SE1 7EH. This study is insured by Guy's & St. Thomas' NHS Foundation Trust under the clinical negligence scheme for trials.

All professional staff involved in the study hold professional indemnity to work within Guy's & St. Thomas' NHS Trust. In the event that you are harmed during the research and this is due to negligence then you may have grounds for legal action for compensation against Guy's & St. Thomas NHS Trust but you may have to pay your legal costs. The normal NHS complaints mechanisms are still available to you.

**Will my taking part in this study be kept confidential?**

All samples collected about you will be anonymised after collection by assigning a study number. We will keep details of your colonoscopy report and the histopathology report, containing your name, age and hospital number. This information will be stored in a locked filing cabinet and will only be accessible by the named investigators in the study. A secure database of samples collected in the study will contain your initials and study number only. After the study is completed, any records where you are personally identifiable will be destroyed.

Any paper will be shredded and disposed of via a secure contractor. Computer records will be destroyed using appropriate data destruction software.

Gut Bacteria in Colorectal cancer

PIS part 2 v3

24th November 2014

**APPENDIX D: PCR Results Summary****Healthy Cohort PCR Results**

<b>Sample ID</b>	<b><i>B. fragilis</i></b>	<b><i>bft</i></b>	<b><i>ubb</i></b>	<b><i>Fusobacterium</i></b>	<b><i>fadA</i></b>
BF012.AC	-	-	-	+	-
BF012.DC	-	-	-	-	-
BF018.AC	+	-	+	-	-
BF018.DC	+	-	+	-	-
BF022.AC	+	-	-	-	-
BF022.DC	+	-	-	-	-
BF031.AC	-	-	-	-	-
BF031.DC	-	-	-	-	-
BF036.AC	-	-	-	+	-
BF036.DC	-	-	-	-	-
BF046.AC	+	-	-	+	-
BF046.DC	+	-	-	+	-
BF051.AC	+	-	-	-	-
BF051.DC	+	-	-	-	-
BF052.AC	-	-	-	-	-
BF052.DC	-	-	-	-	-
BF057.AC	+	-	-	-	-
BF057.DC	+	-	-	-	-
BF063.AC	+	-	-	+	+
BF063.DC	+	-	-	-	-
BF064.AC	+	-	-	-	-
BF064.DC	+	-	-	-	-
BF065.AC	+	-	-	+	+
BF065.DC	+	-	-	+	+
BF074.AC	+	-	-	+	-
BF074.DC	+	-	-	+	-
BF077.AC	-	-	-	-	-
BF077.DC	-	-	-	+	-
BF086.AC	-	-	-	-	-
BF086.DC	+	-	-	+	-
BF088.AC	+	-	-	-	-
BF088.DC	+	-	-	-	-
BF095.AC	+	-	-	-	-
BF095.DC	+	-	-	+	-
BF106.AC	+	-	-	-	-
BF106.DC	+	-	-	-	-
BF111.AC	+	-	-	+	+
BF111.DC	+	-	-	+	+
BF121.AC	+	-	-	+	-

### Chapter 6.1.3 Fusobacterium and F. nucleatum taxonomy

BF121.DC	+	-	-	+	-
BF123.AC	-	-	-	-	-
BF123.DC	-	-	-	-	-
BF124.AC	-	-	-	-	-
BF124.DC	-	-	-	-	-

Sample ID is Patient ID. Biopsy location

AC = ascending colon; DC = descending colon, OT = on tumour, NT = next to tumour, AT = away from tumour

- = PCR negative + = PCR positive

**IBD Cohort PCR Results**

Sample ID	<i>B. fragilis</i>	<i>bft</i>	<i>ubb</i>	<i>Fusobacterium</i>	<i>fadA</i>
BF001.DC	+	-	-	-	-
BF003.AC	+	+	+	-	-
BF003.DC	+	+	-	-	-
BF014.AC	+	-	-	+	+
BF014.DC	+	-	-	-	-
BF015.AC	+	-	-	-	-
BF015.DC	+	-	-	-	-
BF019.OT	+	-	-	-	-
BF019.NT	+	-	+	-	-
BF020.AC	-	-	-	-	-
BF020.DC	-	-	-	-	-
BF033.AC	+	+	+	-	-
BF033.DC	+	+	+	-	-
BF035.AC	+	-	-	-	-
BF041.AC	+	-	-	+	-
BF041.DC	+	-	-	-	-
BF049.OT	-	-	-	-	-
BF049.NT	-	-	-	-	-
BF049.AT	-	-	-	-	-
BF056.AC	-	-	-	-	-
BF056.DC	-	-	-	-	-
BF070.AC	-	-	-	-	-
BF070.DC	-	-	-	-	-
BF084.AC	-	-	-	+	-
BF084.DC	-	-	-	-	-
BF099.AC	+	-	-	-	-
BF099.DC	+	-	-	-	-
BF117.AC	+	-	-	-	-
BF117.DC	-	-	-	-	-

Sample ID is Patient ID. Biopsy location

AC = ascending colon; DC = descending colon, OT = on tumour, NT = next to tumour, AT = away from tumour

- = PCR negative + = PCR positive

**Previous polyps/cancer cohort PCR results**

<b>Sample ID</b>	<b><i>B. fragilis</i></b>	<b><i>bft</i></b>	<b><i>ubb</i></b>	<b><i>Fusobacterium</i></b>	<b><i>fadA</i></b>
BF014.AC	+	-	-	+	+
BF014.DC	+	-	-	-	-
BF016.AC	-	-	-	+	-
BF016.DC	-	-	-	-	-
BF017.AC	-	-	-	-	-
BF017.DC	-	-	-	-	-
BF024.AC	-	-	-	-	-
BF042.AC	+	-	-	-	-
BF042.DC	+	-	-	-	-
BF045.AC	-	-	-	-	-
BF045.DC	+	-	-	+	-
BF059.AC	+	-	+	-	-
BF059.DC	+	-	+	-	-
BF060.AC	-	-	-	-	-
BF060.DC	-	-	-	-	-
BF071.AC	+	-	-	-	-
BF071.DC	+	-	-	-	-
BF076.AC	+	-	-	-	-
BF076.DC	+	-	-	-	-
BF080.AC	+	+	-	-	-
BF080.DC	+	+	-	-	-
BF091.AC	-	-	-	-	-
BF091.DC	-	-	-	-	-
BF097.AC	+	-	-	-	-
BF097.DC	+	-	-	-	-
BF098.AC	-	-	-	-	-
BF098.DC	-	-	-	-	-

Sample ID is Patient ID. Biopsy location

AC = ascending colon; DC = descending colon, OT = on tumour, NT = next to tumour, AT = away from tumour

- = PCR negative + = PCR positive



**New polyps/cancer cohort PCR results**

<b>Sample ID</b>	<b><i>B. fragilis</i></b>	<b><i>bft</i></b>	<b><i>ubb</i></b>	<b><i>Fusobacterium</i></b>	<b><i>fadA</i></b>
BF002.OT	-	-	-	-	-
BF002.NT	-	-	-	-	-
BF002.AT	-	-	-	+	-
BF005.OT	-	-	-	-	-
BF005.NT	-	-	-	-	-
BF006.OT	+	-	-	+	+
BF006.NT	+	-	-	+	+
BF006.AT	+	-	-	+	+
BF007.OT	-	-	-	-	-
BF007.NT	-	-	-	-	-
BF007.AT	-	-	-	-	-
BF008.OT	+	+	-	+	+
BF008.NT	+	+	-	+	+
BF008.AT	+	+	-	+	+
BF009.AC	+	-	-	-	-
BF009.DC	+	-	-	-	-
BF011.AC	+	-	-	+	-
BF011.DC	+	-	-	-	-
BF013.OT	+	-	-	-	-
BF013.AC	+	-	+	+	-
BF019.OT	+	-	-	-	-
BF019.NT	+	-	+	-	-
BF027.OT	+	+	-	-	-
BF027.NT	+	+	-	-	-
BF027.AT	+	+	-	-	-
BF028.OT	+	+	-	-	-
BF028.NT	+	+	-	-	-
BF028.AT	+	+	-	-	-
BF029.OT	+	-	-	-	-
BF029.NT	+	-	-	-	-
BF029.AT	+	-	-	-	-
BF030.AC	-	-	-	-	-
BF030.DC	-	-	-	-	-
BF032.DC	-	-	-	-	-
BF034.OT	+	-	-	-	-
BF034.NT	-	-	-	-	-
BF034.AT	-	-	-	-	-
BF035.AC	+	-	-	-	-
BF041.AC	+	-	-	+	-
BF041.DC	+	-	-	-	-
BF043.AC	+	+	-	-	-

### Chapter 6.1.3 Fusobacterium and F. nucleatum taxonomy

BF044.AC	-	-	-	-	-
BF047.AC	-	-	-	+	-
BF047.DC	-	-	-	+	-
BF048.OT	-	-	-	-	-
BF048.NT	-	-	-	-	-
BF048.AT	-	-	-	-	-
BF049.OT	-	-	-	-	-
BF049.NT	-	-	-	-	-
BF049.AT	-	-	-	-	-
BF054.AC	-	-	-	-	-
BF054.DC	-	-	-	-	-
BF056.AC	-	-	-	-	-
BF056.DC	-	-	-	-	-
BF058.OT	-	-	-	-	-
BF058.AC	-	-	-	-	-
BF061.OT	+	-	+	+	-
BF061.NT	+	-	+	+	-
BF061.AT	+	-	+	-	-
BF071.AC	+	-	-	-	-
BF071.DC	+	-	-	-	-
BF072.OT	-	-	-	-	-
BF072.NT	-	-	-	-	-
BF072.AT	-	-	-	-	-
BF079.NT	+	-	-	-	-
BF079.AT	+	-	-	-	-
BF090.AC	+	-	+	+	+
BF090.DC	+	-	+	+	+
BF100.AC	-	-	-	-	-
BF100.DC	-	-	-	-	-
BF115.OT	-	-	-	-	-
BF115.NT	-	-	-	-	-
BF115.AT	-	-	-	-	-

Sample ID is Patient ID. Biopsy location

AC = ascending colon; DC = descending colon, OT = on tumour, NT = next to tumour, AT = away from tumour

- = PCR negative + = PCR positive

**APPENDIX E****Metagenomic Sequences retrieved from KEGG nucleotide BLAST used in DNA sequence alignments**

KEGG Sample name	MetaHIT sample name	Gene	Sequence position	Sex	Age	Health Status	Nationality
T30082	MH0068	<i>bft-1</i>	scaffold939_15:3969..5186	F	54	Healthy	Danish
T30068	MH0054	<i>bft-2</i>	scaffold81981_3:complement(4090..5307)	M	49	Healthy	Danish
T30112	O2. UC-22	partial <i>bft-1</i>	C4157881_1:380..772	M	44	IBD	Spanish
T30099	MH0085	<i>ubb</i>	scaffold58215_1:complement(450..761)	F	59	Healthy	Danish
T30094	MH0080	<i>ubb</i>	scaffold64380_5:82..393	F	59	Healthy	Danish